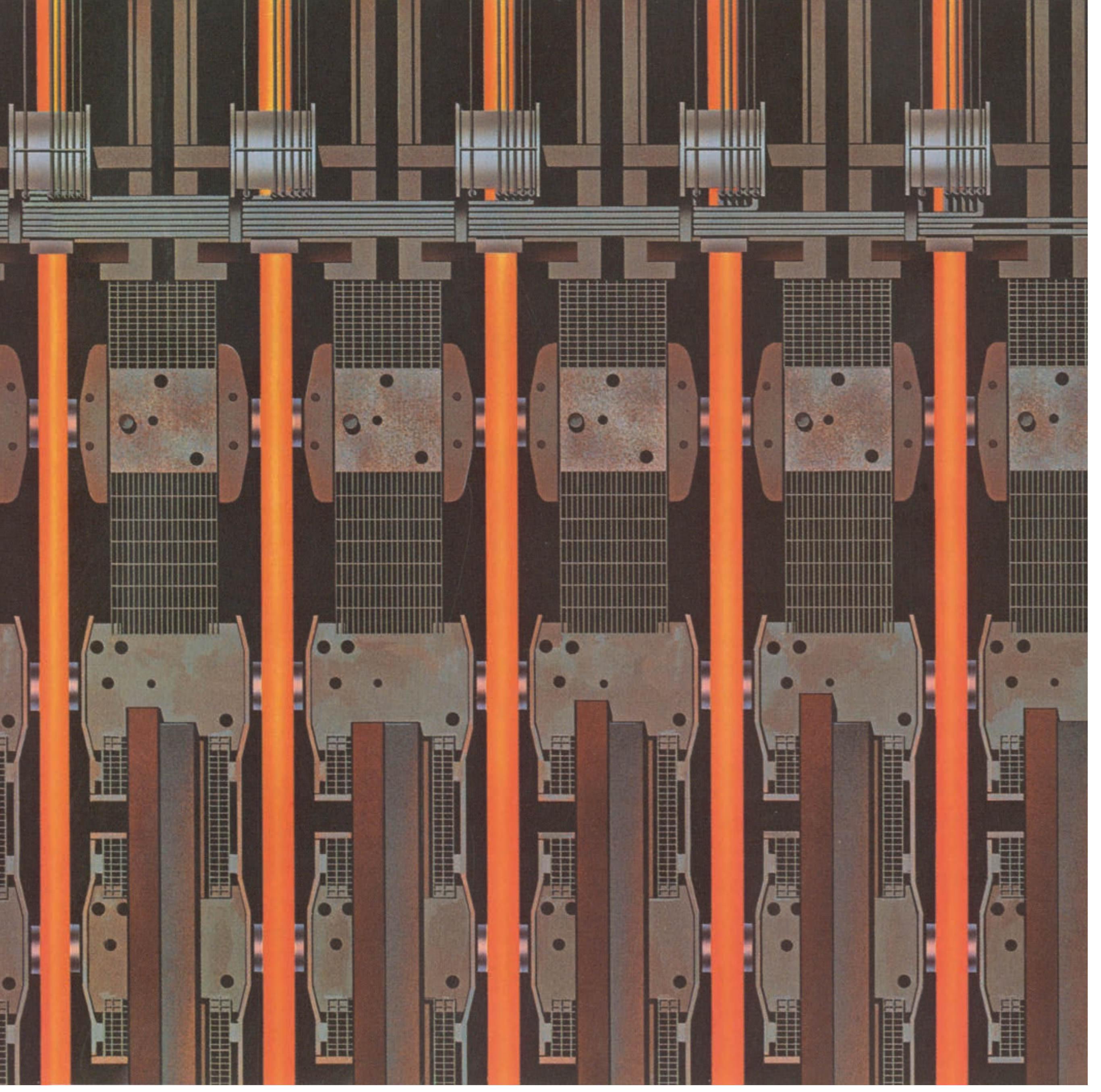
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



MINIACERIAS

Julio 1984

350 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 94
Julio 1984

INVESTIGACION Y CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

Colaboradores de este número:

Asesoramiento y traducción:

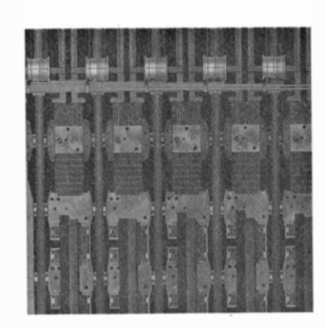
Luis Bou: *Máquinas de Turing* y *Juegos de ordenador*; Montserrat Domingo: *Zonas oceánicas de fractura*; Juan C. Gómez-Fernández: *Endocitosis de proteínas y partículas*; Jordi Jorba: *Miniacerías*; Joandomènec Ros: *Parásitos que cambian el comportamiento de su patrón*; Luis Mas: *El Universo inflacionario*; Antonio Blanco Freijeiro: *Un campamento de cazadores indios durante 20.000 años*; J. Vilardell: *Taller y laboratorio*.

Ciencia y sociedad:

Ramón Margalef y Pedro Molera

Libros:

Luis García Ballester, Javier de Lorenzo, Rafael Sentandreu y Angel Guerra



LA PORTADA

La portada muestra una parte de las cinco palanquillas continuas que emergen de la operación de colada continua en la miniacería de la Raritan Steel Co. instalada en Perth Amboy, Nueva Jersey. En la colada continua moderna, el acero fundido se vacía desde una cuchara a un distribuidor, estructura que se parece a un embudo. De allí, desciende verticalmente hasta unos moldes refrigerados por agua y después a una zona con agua pulverizada donde se enfría. Las palanquillas continuas atraviesan una sección curva ayudadas por unos rodillos propulsores. En la parte superior se aprecia las palanquillas en movimiento horizontal a través de la zona de enfriamiento (hacia abajo según se ve la figura). Las palanquillas se cortan a una longitud de 15 metros mediante soplete, los tubos están asociados con el sistema de corte. Cuando las palanquillas se han enfriado y solidificado, se almacenan para su envío posterior a la planta de laminación de varillas. Las miniacerías se distinguen de las acerías por emplear chatarra como materia prima (véase "Miniacerías", por Jack Robert Miller, en este mismo número).

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
08029-Barcelona (España)
Teléfono 322 05 51 ext. 33-37

Condiciones de suscripción:
España:
Un año (12 números): 3850 pesetas
Extranjero:
Un año (12 números): 33 U.S. $
Ejemplar atrasado ordinario:
350 pesetas
Ejemplar atrasado extraordinario:
450 pesetas

Distribución para España:
Distribuciones de Enlace, S. A.
Ausias March, 49 - 08010-Barcelona

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - 08029-Barcelona

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137 - 28030-Madrid
Teléfono 430 84 81
Cataluña:
Lourdes Padrós
Manuel Girona, 61, 2.º - 08034-Barcelona
Teléfono 204 45 83

Controlado por O.J.D.

PROCEDENCIA DE LAS ILUSTRACIONES

Pintura de la portada de Ted Lodigensky

Página	Fuente
9-19	Edward Bell
20-21	Kathleen Crane, Observatorio Geológico Lamont-Doherty (*arriba*); Andrew Tomko (*abajo*)
22	Andrew Tomko
23	William F. Haxby, Observatorio Geológico Lamont-Doherty
24-27	Andrew Tomko
28	Enrico Bonatti, Observatorio Geológico Lamont-Doherty (*arriba*); Dee Breger, Observatorio Geológico Lamont-Doherty (*abajo*)
30-32	Andrew Tomko
35	M. M. Perry y A. B. Gilbert, Centro de Investigación Avícola del Consejo de Investigación Agrícola de Edimburgo (*arriba*); Hans J. Geuze, Universidad estatal de Utrecht (*abajo*)
36-38	Bunji Tagawa
39	Frederick R. Maxfield, Hospital Clínico de la Universidad de Nueva York
40-41	Bunji Tagawa
47	Barrie Rokeach
48-54	Jerome Kuhl
59-61	Tom Prentiss
62	Janice Moore, cortesía de *The Journal of Parasitology*
63-64	Ilil Arbel
67-76	Ian Worpole
81	William C. Johnson, Universidad de Pittsburgh
82-83	Ilil Arbel
84	J. M. Adovasio, Universidad de Pittsburgh
87-95	José A. Cabezas, Pedro Calvo y Ricardo Génova
97-101	Ilil Arbel
102-107	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Pedro IV, 160 - Barcelona-5
Fotocromos reproducidos por GINSA, S.A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Nápoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Los autores

JOHN E. HOPCROFT ("Máquinas de Turing") es profesor de ciencias de cómputo de la Universidad de Cornell. Formado en la Universidad de Seattle, se doctoró por Stanford en 1964. Seguidamente ingresó en el claustro de la Universidad de Princeton, donde impartió durante tres años, para luego trasladarse a Cornell; allí ha permanecido ininterrumpidamente, a excepción de un año, en que fue profesor de la Universidad de Stanford.

ENRICO BONATTI y KATHLEEN CRANE ("Zonas oceánicas de fractura") son investigadores del Observatorio Geológico Lamont-Doherty de la Universidad de Columbia, que colaboran en los estudios descritos en su artículo. Bonatti, italiano de nacimiento e investigador científico del observatorio, resume su biografía académica: "Me licencié en geología por la Universidad de Pisa y por la Escuela Normal Superior de Pisa. En 1960 fui a los Estados Unidos con una beca Fulbright. Estuve un año en la Universidad de Yale y cuatro en la Institución Scripps de Oceanografía. Durante varios años enseñé geología marina en la Universidad de Miami; en 1976 me incorporé al Lamont". Crane se licenció por la Universidad estatal de Oregon y se doctoró por la Institución Scripps de Oceanografía.

ALICE DAUTRY-VARSAT y HARVEY F. LODISH ("Endocitosis de proteínas y partículas") son biólogos moleculares que trabajaron juntos sobre la materia de su artículo cuando se encontraban en el Instituto de Tecnología de Massachusetts (MIT). Dautry-Varsat se licenció en física del estado sólido por la Universidad de París, tras lo cual decidió especializarse en biología molecular. Tras una estancia en el laboratorio de Biología Molecular del Consejo de Investigación Médica de Inglaterra y en el MIT, recaló en el Instituto Pasteur. Lodish es profesor de biología del MIT. Licenciado por el Kenyon College, se recibió de doctor por la Universidad Rockefeller. Ingresó en el claustro docente del MIT en 1968. Este mismo año trasladará su laboratorio al Instituto Whitehead de Investigación Biomédica, del MIT.

JACK ROBERT MILLER ("Miniacerías"), ingeniero consultor y economista, ha estado vinculado a la industria del hierro y del acero durante más de cuarenta años. Licenciado en ingeniería eléctrica por el Instituto Union Cooper, cursó, también, economía en la Universidad de Nueva York. Además de trabajar en varias e importantes acerías norteamericanas, ha colaborado con el Centro de las Naciones Unidas para el Desarrollo Industrial, el Banco Mundial, el Banco Interamericano de Desarrollo y el Instituto Battelle. "En 40 años de asesor, nos explica, he participado en cerca de 50 proyectos de ultramar y al menos otros tantos en los Estados Unidos. En 1976 tuve que jubilarme por razones de salud. Después de seis años de paro involuntario, vuelvo ahora a trabajar. En 1983 establecí de nuevo una asesoría en Sant Petersburg, Florida, pero con una diferencia: somos un grupo de jubilados que no deseamos estarlo."

JANICE MOORE ("Parásitos que cambian el comportamiento de su patrón") se halla adscrito a la Universidad estatal de Colorado. "Siendo estudiante de la Universidad de Rice, confiesa, asistí a un curso de parasitología por razones que he olvidado, y lo terminé con una permanente sensación de maravilla ante tal diversidad y variedad de adaptaciones. Acabé mi tesis de licenciatura en comportamiento de insectos en la Universidad de Texas en Austin, y empecé a estudiar organismos parásitos. Por entonces el campo de la ecología de los parásitos se consideraba interdisciplinar, y se instó a que eligiera un proyecto más en boga. No cedí, y cambié dos veces de facultad y desempeñé diversos trabajos editoriales y técnicos. Conocí luego a varios miembros de la facultad de biología de la Universidad de Nuevo México; y con su estímulo penetré en un mundo en el que desde entonces no he dejado de descubrir fenómenos fascinantes."

ALAN H. GUTH y PAUL J. STEINHARDT ("El Universo inflacionario") son físicos con un interés común en la historia de los primeros instantes del universo. Guth cursó sus estudios en el MIT, por el que se doctoró en 1972. "Trabajé, comenta, en las universidades de Princeton, Columbia y Cornell, y en el Centro del Acelerador Lineal de Stanford (SLAC). Habiendo tratado durante todo ese tiempo problemas matemáticos bastante abstractos sobre teoría de partículas elementales, no conocía de los desarrollos cosmológicos más de lo que alcanzan los profanos. Sin embargo, encontrándome en Cornell, mi colega Henry Tye me convenció para que estudiara con él la producción de monopolos magnéticos en el universo primitivo. Ahí cambió de rumbo mi carrera. Seguí por un curso más mis trabajos en el SLAC y poco después volví al MIT en calidad de profesor adjunto de física, cargo que sigo ocupando." Steinhardt cursó el primer ciclo de ciencias en el Instituto de Tecnología de California. Se licenció (1975) y doctoró en física (1978) por la Universidad de Harvard. De 1979 a 1981 disfrutó de una beca de esa institución. En 1981 se trasladó a la Universidad de Pennsylvania.

J. M. ADOVASIO Y R. C. CARLISLE ("Un campamento de cazadores indios durante 20.000 años") son miembros del departamento de antropología de la Universidad de Pittsburgh. Adovasio es catedrático de antropología y director del departamento. Licenciado en Letras por la Universidad de Arizona, se doctoró por la de Utah, en 1970. Su especialidad científica es la excavación de yacimientos arqueológicos cerrados, como cuevas y abrigos, y el estudio de las técnicas por las que en tiempos antiguos se hacían cestos. En sus actividades extracientíficas Adovasio es motociclista altamente cualificado y levantador de pesos en competiciones. Carlisle es ayudante de Adovasio, ayudante también de investigación, instructor de antropología y redactor del Programa de Gerencia de los Recursos Culturales, de la Universidad. En la actualidad trabaja en su doctorado en Pittsburgh.

JOSE A. CABEZAS y PEDRO CALVO ("Gangliósidos") son catedráticos de bioquímica de la Universidad de Salamanca, el primero, y de la de León, el segundo. Cabezas se doctoró con premio extraordinario en Madrid en 1956. Completó su formación en las universidades de París, Colonia y Pensylvania. Miembro fundador de la Sociedad de Bioquímica Española y de la European Society for Comparative Biochemistry and Physiology, pertenece también a otras instituciones europeas y americanas. Colabora en programas de investigación conjunta con el Instituto Pasteur de París y con otros centros de Inglaterra, Alemania y Estados Unidos. Calvo se formó en la Universidad de Salamanca. Entre sus estancias en el extranjero recuerda con agrado los dos años transcurridos en la Facultad de Medicina de Virginia.

Hace...

José M.ª López Piñero

... ciento cincuenta años

Inició su dedicación a la anatomía José M.ª Gómez Alamá, típica figura de las llamadas "generaciones intermedias" que posibilitaron la recuperación del cultivo de las disciplinas biomédicas en la España de las décadas centrales del siglo XIX.

La biografía de Gómez Alamá es la de un modesto profesor universitario que consagró su vida entera a la enseñanza de las ciencias morfológicas, por lo que puede servir de contrapunto a la de la mayoría de las celebridades médicas españolas de su tiempo, cuyo relieve correspondió al ejercicio profesional o a la actividad política. Nacido en Valencia en 1815, cursó en esta ciudad todos los estudios, incluidos los de medicina. Se interesó por la anatomía al estudiarla con Vicente Llobet, catedrático entonces de la disciplina en la Universidad de Valencia e infatigable disector. Al terminar el curso en 1834, sabemos, por testimonio de su discípulo Nicolás Ferrer Julve, que Llobet "le recomendó al disector anatómico para fomentar sus aptitudes". De esta forma, la ejemplar vocación de Gómez Alamá tuvo su raíz en un ambiente en el que se enseñaba anatomía directamente sobre el cadáver, de acuerdo con la mejor tradición del Colegio de Cirugía de Cádiz, de donde Llobet procedía. Es interesante tener en cuenta esta tardía repercusión de la gran reforma encabezada por Virgili y del excelente nivel del saber anatómico español de la Ilustración en un período de profunda decadencia del cultivo de las ciencias médicas básicas en nuestro país.

El interés que supo despertar Llobet en el joven estudiante no fue algo momentáneo: en 1835, "el mismo claustro dispuso... se le facilitaran cadáveres y todos los instrumentos y enseres necesarios fuera de las horas de enseñanza pública, para que se perfeccionara más y más en un ramo tan importante de los estudios médicos". Como hemos adelantado, el resto de la vida de Gómez Alamá estuvo dedicada exclusivamente a la anatomía. En 1836 fue designado ayudante de disector, y tras varios años de profesor sustituto y agregado, en los que tuvo que enseñar las más diversas asignaturas, consiguió en 1848 ser nombrado catedrático de anatomía de la misma Universidad de Valencia. Los méritos principales que se adujeron en dicho nombramiento fueron "los brillantes y especiales servicios" que realizó como director de trabajos anatómicos de la institución. A este respecto, merece ser recogido el contraste de su figura con la de Agapito Zuriaga, entonces titular de la otra cátedra de anatomía de Valencia, cargo que utilizó, igual que tantos otros profesores españoles de su tiempo y el nuestro, como plataforma de una buena clientela particular de cirujano. Un detalle, carente de importancia en sí mismo, refleja de modo muy significativo el abismo que separaba los dos tipos de profesor que con tanta propiedad representan ambos. En un elogio de Zuriaga publicado por su sucesor en la cátedra, se afirmaba que la colección de piezas anatómicas artificiales y naturales del gabinete de anatomía de la Facultad de Medicina de Valencia procedían de un viaje que éste había realizado a Francia, Bélgica e Inglaterra en 1848, comisionado para la adquisición de material científico. En el número siguiente de la misma revista médica que había publicado tal elogio, apareció una rectificación en la que, entre otras cosas, se decía: "la mayor parte de (las piezas anatómicas) que se admiran en el mencionado gabinete son debidas a la laboriosidad y destreza del... doctor don José Gómez... Este digno profesor, auxiliado por dos ayudantes, fue el que tomó la iniciativa en el asunto que se trata y llevó a cabo, en tiempo de vacaciones, su laboriosa empresa".

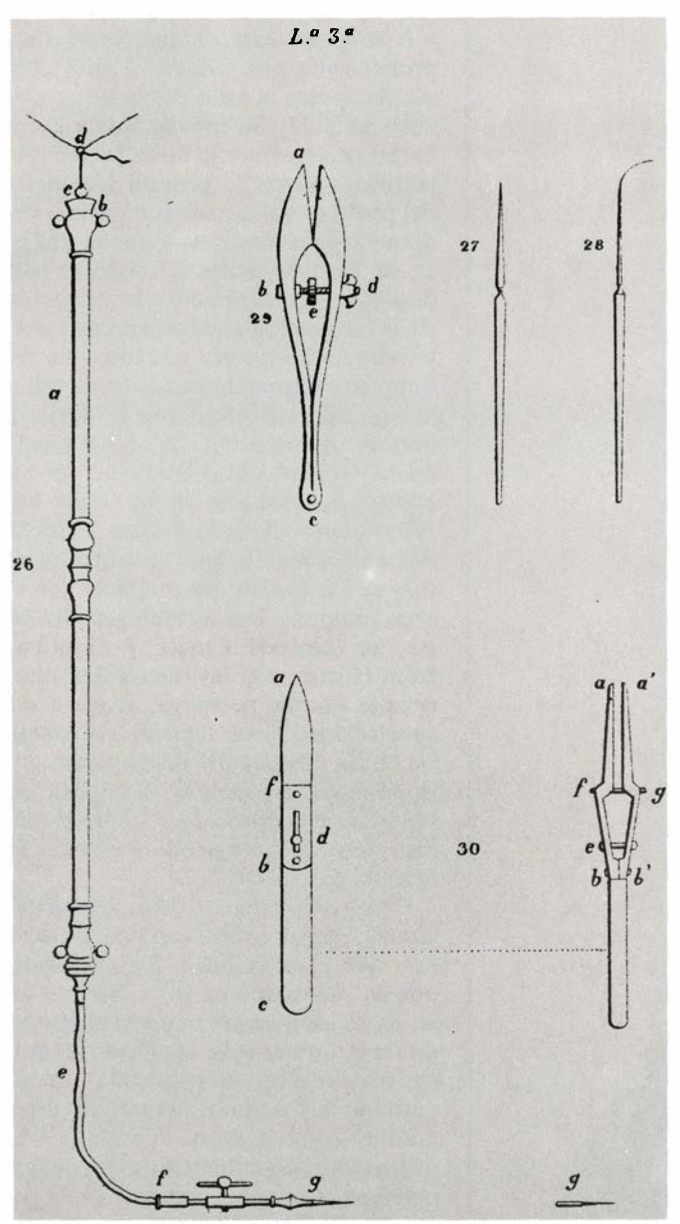
Durante el cuarto de siglo de permanencia en la cátedra, Gómez Alamá fue estrictamente fiel a su dedicación a la enseñanza, como lo había sido en la primera parte de su vida docente. El historiador de la medicina Luis Comenge, que fue discípulo suyo, nos ha dejado un expresivo testimonio: "Don José María Gómez Alamá perteneció a los preceptores que sin estar adornados de arrebatadora elocuencia... esculpía en la mente del escolar las enseñanzas áridas de la asignatura, infundía el gusto hacia tales estudios, y en su cátedra no trató nunca de otra cosa que no fuese la exposición oral o la demostración práctica de las lecciones...". En el mismo sentido hablan sus únicas producciones científicas, consagradas exclusivamente a fines didácticos, y aparecidas casi al final de su vida, como el producto de su experiencia de profesor. También en este punto contrasta su figura con la de Zuriaga, que redactó su *Compendio de anatomía* siendo estudiante y que se sirvió de él como escalón hacia la cátedra. Las publicaciones de Gómez Alamá fueron dos ediciones de un manual de anatomía descriptiva y de un *Arte de disecar*, y un *Discurso sobre la importancia de la anatomía humana*, escrito "con el fin de animar a sus discípulos a no abandonar el estudio de una ciencia tan interesante y transcendental". Todas ellas aparecieron entre 1867 y 1872. Poco más tarde, en 1874, fallecía su autor en la misma ciudad de Valencia.

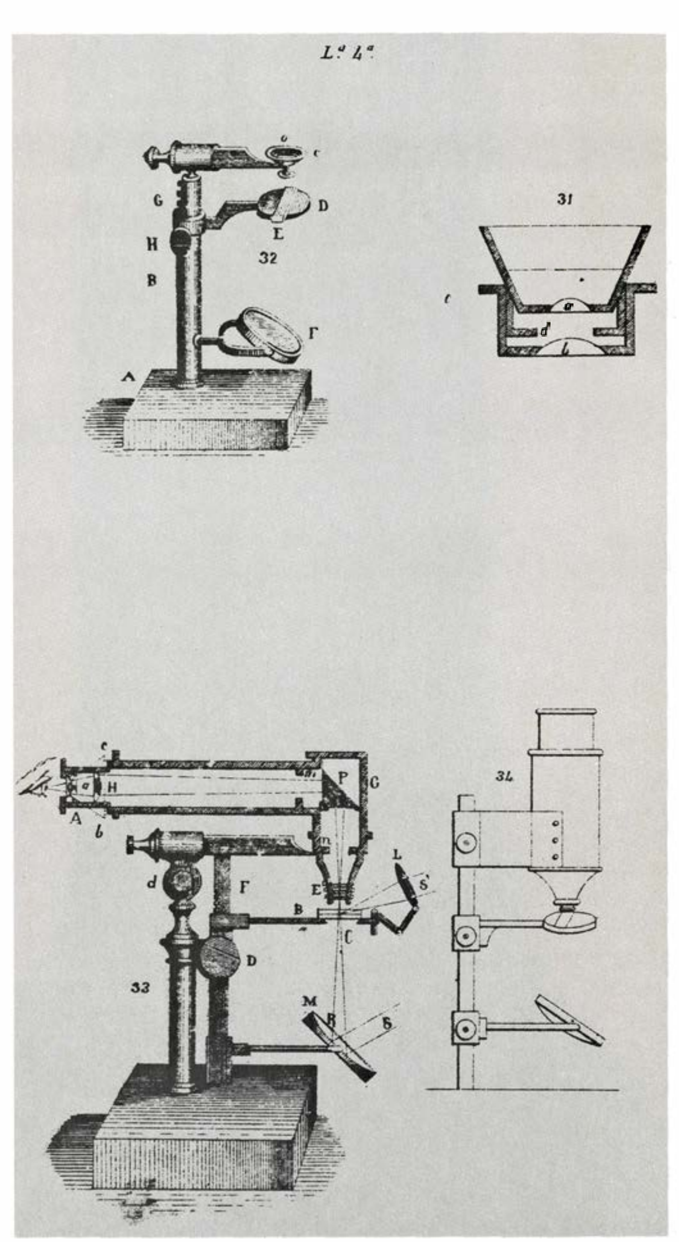
La primera edición de su manual apareció en dos volúmenes bajo el título de *Compendio de Anatomía descriptiva y elementos de la general con nociones de anatomía microscópica* (1867-1868). Se inicia con unas "nociones preliminares", destinadas principalmente a exponer los supuestos generales que sirven de base al estilo anatómico descriptivo utilizado. El profesor valenciano maneja sin pretensiones los esquemas didácticos de la anatomía de su tiempo, muy especialmente de la francesa; su *idea descriptiva* es estática o arquitectural, delimita conceptualmente las *partes anatómicas* con el eclecticismo habitual que combinaba los criterios funcional y morfológico, y en su *método descriptivo* utiliza el cubo como figura de referencia. Aplica estos principios al estudio de las diferentes partes anatómicas, según el orden siguiente: huesos, articulaciones y músculos, aponeurosis y tegumentos, "esplagnolegia", vasos, órganos de los sentidos y sistema nervioso. La única cualidad sobresaliente es la claridad de la redacción y de la sistematización del contenido, para contribuir a la cual se inserta, en una obra que carece desgraciadamente de iconografía, una amplia serie de excelentes cuadros sinópticos.

Gómez Alamá se ocupa también de la "estructura y composición" del cuer-

po humano, epígrafe en el que incluye unas breves noticias bioquímicas y un resumen de la teoría celular. Tras hablar de los tejidos y de las fibras conectiva, muscular y nerviosa, Gómez Alamá acepta la célula como "elemento anatómico inmediato". A pesar de la modestia de la información que proporciona sobre el tema, la contribución no carece de importancia, por tratarse de la primera vez que un texto español de anatomía destinado a la enseñanza incorporó sin restricciones tan decisiva novedad doctrinal. Anotemos también que estudia con cierta amplitud los diferentes tejidos de acuerdo con el más estricto criterio celularista y con notable acúmulo de datos microscópicos.

De menos interés es la parte que dedica a la embriología, reducida a breves notas descriptivas de la génesis de las principales formaciones. Aunque Gómez Alamá no ofrece ninguna exposición general de las teorías embriológicas que le parecían más aceptables, es evidente que, como sucedía a otros anatomistas españoles de estos años, su





Láminas del **Arte de disecar** *(2.ª ed., 1872), de José M.ª Gómez Alamá, en la que aparecen varios de los instrumentos que utilizaba para la enseñanza de la anatomía.* **26,** *"Aparato de inyección mercurial de los vasos linfáticos por medio de la gravedad".* **27** *y* **28,** *"Agujas de disección con mango fijo, para anatomía microscópica".* **29,** *"Microtomo de Strauss-Durkheim, especie de pinza cortante destinada a obtener por incisión láminas muy delgadas de las partes anatómicas de mediana consistencia, destinadas a la observación microscópica".* **30,** *"Discotomo de Valentin. Es un cuchillo de dos hojas paralelas por sus planos que pueden aproximarse a voluntad y que sirve para cortar láminas muy delgadas de los parénquimas blandos, para someterlas al microscopio".* **31,** *"Doblete. Aparatito compuesto de dos lentes plano-convexas desiguales en latitud y colocadas de modo que ambas tengan su cara plana dirigida hacia el objeto que se inspecciona".* **32,** *"Microscopio simple de Raspail. En el centro de la capsulita C se halla engastada una lente 0 convergente o un doblete o un triplete que pueden cambiarse para variar el aumento".* **33,** *"Esquema del microscopio compuesto de Amici o de Chevalier. Si el objeto C, colocado entre dos láminas de vidrio, es trasparente, se le ilumina por medio de la luz reflejada por el espejo cóncavo M. Pero si es opaco, se le alumbra con la luz incidente recogida y convenientemente dirigida por la lente* **L***".* **34,** *"Hay también microscopios compuestos de tubo recto vertical, pero son incómodos, porque obligan al observador a adoptar una posición violenta".*

base ideológica son los estudios del francés Serres.

En 1872 apareció una segunda edición, con el título algo distinto de *Tratado elemental de Anatomía humana descriptiva, general y microscópica*, aunque las diferencias de contenido son de poca entidad. Al final del segundo tomo, se incluye un *Discurso sobre la importancia de la anatomía humana en sus relaciones con las artes, las ciencias y la religión*, texto de una conferencia pronunciada por Gómez Alamá ante sus discípulos al final del curso académico 1871-72. Su interés reside en el hecho de constituir la única fuente que permite conocer la postura ideológica del profesor valenciano por encima del desnudo esquematismo a que se obligó en su libro de texto. El discurso está destinado expresamente a la refutación de la antropología evolucionista y materialista. El espectro temático que refleja es sorprendentemente amplio, aunque en casi todos ellos se advierte que la información es de segunda mano. Gómez Alamá defiende, por supuesto, la existencia de un "reino humano" dentro de la naturaleza, la fijeza de las especies biológicas y la identificación de los fenómenos psíquicos con el alma inmortal. Los morfólogos idealistas, en especial Cuvier y Geoffroy Saint-Hilaire, son las principales autoridades en que se apoya. Junto a sus aportaciones, tiene también en cuenta las de las cabezas del pensamiento antropológico español de la época de orientación conservadora, como el médico José de Letamendi y el filósofo Manuel de Uribarri.

Como complemento de su *Anatomía*, Gómez Alamá publicó un *Arte de disecar*, que tuvo también dos ediciones, una en 1870 y otra en 1872. Resume en él, en un estilo coherente con el utilizado en el *Compendio*, las técnicas clásicas para realizar preparaciones anatómicas de los distintos aparatos y sistemas, así como la forma "de someter al microscopio los tejidos con el fin de inspeccionar su textura íntima". En la segunda edición, añadió algunos capítulos nuevos, el más interesante de los cuales es el titulado "Noción y uso del microscopio". Aunque hay una notable diferencia entre la viveza de muchas de sus recomendaciones relativas a las preparaciones macroscópicas y la sequedad de las correspondientes a la micrografía, subraya "que la anatomía microscópica excita con justo motivo la curiosidad, y debe generalizarse pronto, puesto que está prestando grandes servicios a la ciencia".

Máquinas de Turing

En su fundamento lógico, todo ordenador digital incorpora en sí uno de estos dispositivos conceptuales ideados por el matemático inglés A. M. Turing. Las máquinas demarcan los límites de la computabilidad

John E. Hopcroft

En 1900, el preeminente matemático David Hilbert presentó al Congreso Internacional de Matemáticos, que se celebraba en París, una relación de problemas no resueltos, que supusieron para el mundo de los expertos un reto y un estímulo. El problema segundo de aquella lista consistía en demostrar que los axiomas de la aritmética ordinaria eran consistentes (coherentes) entre sí. El trabajo de Hilbert sobre la cuestión le llevó a proponer un problema más general: el *Entscheidungsproblem,* o problema de la decisión. El *Entscheidungsproblem* iba a descubrir un método general para determinar si una fórmula de la lógica formal podía o no satisfacerse, o declararse verdadera. Treinta y seis años hubieron de transcurrir hasta su resolución completa, que señaló en las matemáticas un giro tan extraordinario como inesperado. En la Universidad de Cambridge, un joven matemático, miembro del King's College, de nombre Alan Mathison Turing, había llegado a familiarizarse con el problema gracias a una serie de lecciones impartidas por M. H. A. Newman. Turing reflexionó sobre el problema durante los largos paseos que por las tardes solía dar por la campiña inglesa. Fue tras uno de esos paseos cuando le llegó la respuesta. El problema de Hilbert era imposible de resolver.

La publicación en la que Turing anunció su resultado ha tenido una significación y trascendencia que rebasan con mucho el problema a que inmediatamente se dirigía. Al atacar el problema de Hilbert, Turing se vio forzado a plantearse cómo dar al concepto de método una definición precisa. A partir de la idea intuitiva de que un método es un algoritmo –un procedimiento que puede ser ejecutado mecánicamente, sin intervención creativa alguna–, Turing hizo ver cómo esta idea puede refinarse y convertirse en un modelo detallado del proceso de computación, en el cual un algoritmo cualquiera es descompuesto en una secuencia de pasos atómicos simples. El modelo computacional resultante es una construcción lógica conocida por máquina de Turing.

La forma más sencilla de describir la máquina de Turing es mediante elementos mecánicos, tales como ruedas, cintas perforadas y un sensor capaz de desplazarse hacia adelante y atrás, a lo largo de la cinta. Si bien toda esta maquinaria no es imprescindible, pues en su nivel fundamental la máquina de Turing no es sino la concreción material de un método de razonamiento matemático, la eliminación total de la analogía mecánica podría inducir a confusiones. Tal analogía le resultaba sugerente al propio Turing, que fue uno de los pioneros en el desarrollo del ordenador digital. Además, los méritos de la máquina de Turing en tanto que instrumento conceptual son hoy proclamados por las ciencias de cómputo con, por lo menos, tanta fuerza como pueda hacerlo la lógica. En las ciencias de cómputo su importancia es fundamental: si a la máquina de Turing se le da tiempo suficiente, que será mucho, pero finito, podrá llevar a cabo cualquier cómputo que pueda realizar un moderno ordenador digital, por muy potente que sea.

Esta capacidad universal de la máquina de Turing no implica que pueda ser utilizada de modo práctico como ordenador. Los ordenadores reales pueden funcionar a velocidad enormemente mayor que la máquina de Turing, porque al diseñarlos se sacrifica deliberadamente la claridad de funcionamiento en aras de mayor eficiencia y velocidad. No obstante, en el estudio teórico de la capacidad última que para la resolución de problemas pueda tener el ordenador real, la máquina de Turing se ha hecho indispensable. La máquina de Turing ha permitido, por ejemplo, demostrar que además del problema de Hilbert hay otros muchos que no pueden ser resueltos, por muy potente y veloz que sea el ordenador utilizado para ello.

Funcionamiento

¿Qué es y cómo funciona una máquina de Turing? Andrew Hodges, en una reciente biografía de Turing, la compara a una máquina de escribir ordinaria. Lo mismo que la máquina de escribir, la de Turing dispone de un cabezal móvil para la impresión, que graba sobre una superficie adecuada símbolos individuales y discretos tomados de un alfabeto finito. Para simplificar los movimientos del cabezal, se supone que la superficie de impresión es una cinta dividida en segmentos o cuadrículas discretas. Así pues, el cabezal de escritura de la máquina de Turing tan sólo tiene que moverse en una dimensión, hacia la derecha o hacia la izquierda, y las acciones de la máquina no tienen necesidad de entrar en complicaciones tan irrelevantes como la longitud de la línea impresa o el interlineado. En cada cuadro de la cinta tan sólo puede grabarse un símbolo, pero no hay limitación sobre la longitud de la cinta, ni, consiguientemente, sobre la longitud de la cadena de caracteres que pueden quedar registrados en ella.

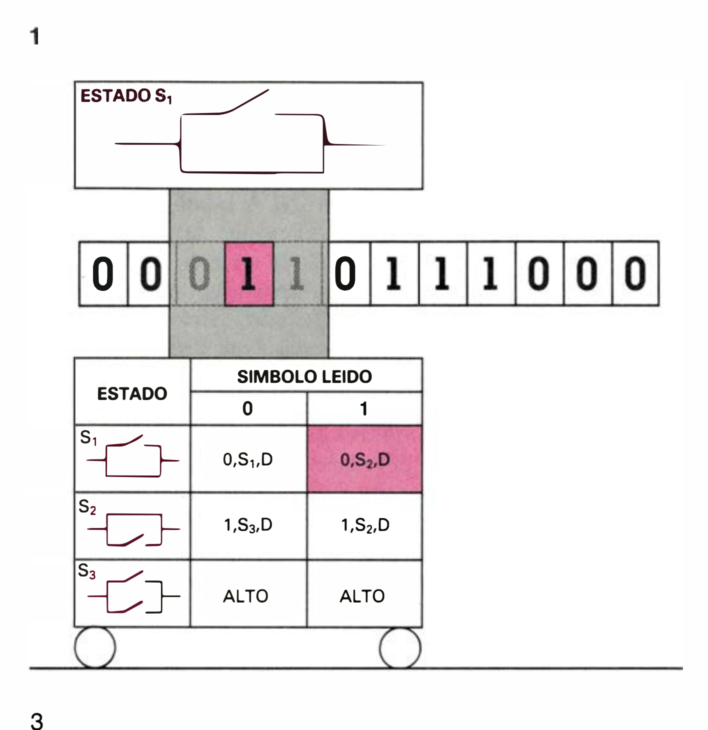
El cabezal móvil de la máquina de Turing puede realizar, además de la impresión, otras dos funciones. Al igual que muchas de las máquinas de escribir fabricadas durante el decenio pasado, puede borrar, o suprimir, uno a uno, los símbolos grabados en la superficie de escritura. Y a diferencia de la máquina de escribir, el cabezal puede también "leer", esto es, reconocer el contenido simbólico de cada cuadro de la cinta, a razón de uno cada vez. De esta forma, los símbolos de la cinta pueden ingresar con carácter de datos en la má-

quina e influir en la determinación de su acción subsiguiente.

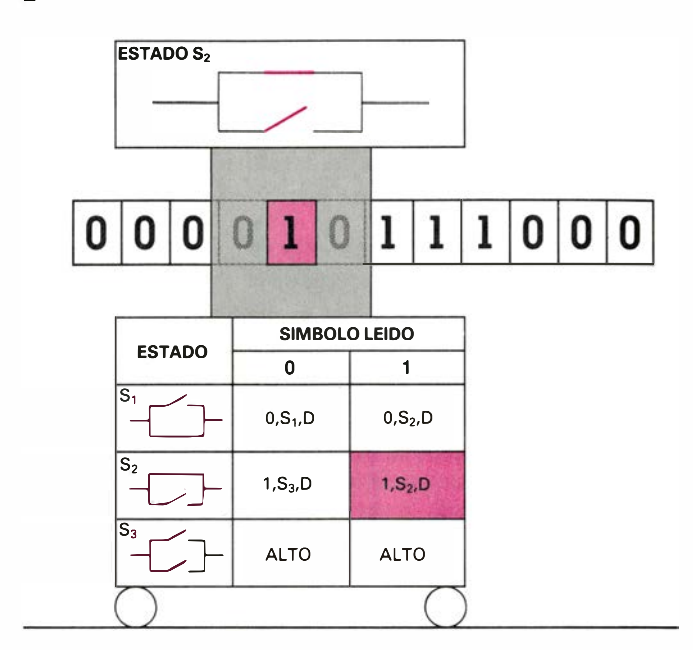
Una máquina de escribir puede adoptar en cada caso uno de varios "estados", o modos de funcionamiento en el transcurso de su actividad. En estado "minúsculas" imprime letras minúsculas o cifras, mientras que en el estado "mayúsculas" imprime éstas, o signos especiales. Análogamente, una máquina de Turing puede ir adoptando uno de entre un número finito de estados. Presumiblemente, cada estado constituye una configuración o disposición distinta de la máquina, pero puesto que la máquina de Turing es un artefacto relativamente abstracto, por lo común no se hace ningún esfuerzo para dar descripciones más concretas y mecánicas de sus estados. Será suficiente describir cada estado de la máquina a partir de los efectos que produzca sobre la actividad de la misma.

La actividad de la máquina de Turing

1

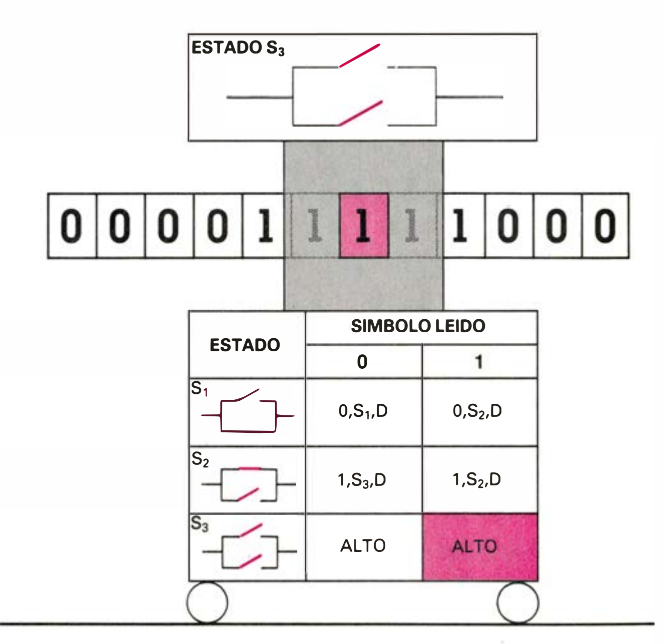


2



3

4



1. LA SUMA DE 2 Y 3 se realiza en cuatro pasos con una máquina de Turing. Cada sumando está representado en notación unaria sobre una cinta, en forma de secuencia de "unos" consecutivos, encerrados entre ceros por ambos extremos. La máquina puede inspeccionar uno por uno los contenidos de las cuadrículas de la cinta (*en color*), yendo y viniendo a izquierda o derecha de ella, en una serie de pasos discretos. La tarea de la máquina consiste en generar una hilera de cinco "unos" consecutivos y detenerse después. La tabla de instrucciones, que vemos bajo cada una de las máquinas del diagrama, es un conjunto fijo de movimientos que cubre todas las condiciones iniciales posibles y que da un procedimiento para sumar dos números cualesquiera. Obedeciendo las instrucciones, la máquina suprime el cero que separa dos hileras de unos y hace avanzar un paso hacia la derecha a la hilera de la izquierda, empalmándola así con la hilera derecha. El número de instrucciones iniciales previstas en la tabla debe ser suficiente para atender todas las contingencias que pudieran presentarse en la cinta; tal número puede incrementarse aumentando el número de estados internos, o configuraciones. Para cada posible combinación de símbolo de la cinta y de estado de la máquina, la tabla de instrucciones debe, o bien detener la máquina, o bien especificar tres variables: primero, dar el símbolo que ha de quedar en la casilla en ese momento inspeccionada; segundo, especificar el estado que la máquina ha de adoptar antes de examinar otra cuadrícula; y tercero, indicar si la máquina ha de desplazarse una cuadrícula hacia la derecha o la izquierda de la cinta.

se compone enteramente de pasos instantáneos y discretos; cada paso está determinado por dos condiciones iniciales: el estado presente de la máquina y el símbolo que está situado en el cuadro de la cinta que en ese momento se esté inspeccionando. Dado un cierto par de condiciones iniciales, la máquina recibe para su próximo paso una instrucción compuesta por tres secciones. La primera parte de la instrucción designa el símbolo que la máquina ha de depositar en el cuadro sometido a examen. Por ejemplo, si la instrucción especifica que en el cuadro ha de quedar el símbolo 1, la máquina imprimirá este símbolo si el cuadro está en blanco, dejará el símbolo intacto si hay un 1 ocupándolo y, si hubiera otro símbolo distinto de 1, lo borraría y reemplazaría por un 1.

La segunda parte de la instrucción especifica cuál será el siguiente estado de la máquina. Al igual que en la especificación del símbolo en la primera parte de la instrucción, la designación de un estado puede requerir que la máquina cambie su estado, o que conserve el que ya tiene. La tercera parte de la instrucción especifica si el cabezal impresor ha de inspeccionar en la cinta un cuadro hacia la izquierda o hacia la derecha.

Podemos abreviar la instrucción completa mencionando los tres valores de sus variables –el símbolo de la cinta, el estado de la máquina y el desplazamiento del cabezal de impresión– en un orden prefijado. Por ejemplo, si un par de condiciones iniciales han de hacer que la máquina deposite el símbolo 1 en el cuadro sometido a inspección, hacer que la máquina adopte el estado S_2 y que desplace un cuadro hacia la izquierda el cabezal de impresión, la instrucción podría abreviarse $(1,S_2,I)$.

La mejor manera de comprender cómo funciona una máquina de Turing es tratar de construir una. En este contexto, construir una máquina de Turing significa construir una tabla de instrucciones que especifiquen la actuación de la máquina en cada posible par formado por un estado y un símbolo de la cinta. En la práctica, la construcción de la tabla conlleva además dotar a la máquina de un número suficiente de estados, que la permitan realizar la tarea que se le encomienda.

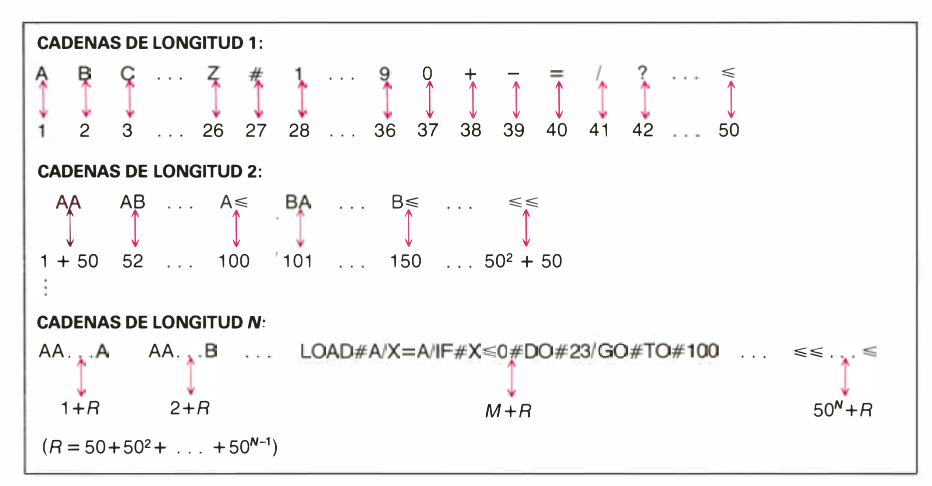
La máquina de sumar

Fijémonos en cómo podríamos proceder para diseñar una máquina que realice la suma de dos números y se detenga. Es costumbre, aunque en modo alguno es esencial, hacer que en la cinta se impriman solamente dos símbolos, por ejemplo, 0 y 1. Cualquier número N dado puede ser entonces representado en la cinta por una serie de N símbolos 1 consecutivos. Si es necesario imprimir en la cinta dos números, el M y el N, podemos representarlos por una primera serie de M "unos", seguida a la derecha de un 0, y seguida consecutivamente a su derecha de otra serie de N "unos". Supongamos que la máquina de Turing se encuentre en un estado S_1 inicial y que su cabezal esté inspeccionando el primero desde la izquierda de los 1 de la serie que expresa a M. Para construir una máquina de Turing capaz de sumar, lo que se requiere es generar una tabla de instrucciones capaz de lograr que la máquina imprima una serie de $M + N$ "unos" consecutivos y después se detenga.

Una forma sencilla de ejecutar con éxito esta tarea es llevar un cuadro hacia la derecha a toda la serie de "unos" correspondientes a M, que es, de las dos series, la situada más a la izquierda. Si tal desplazamiento es correctamente realizado, el 0 dejará de separar las dos cadenas de "unos"; la serie única de "unos" resultante tendrá longitud $M + N$. A la máquina de Turing no le es posible desplazar de un solo golpe toda la serie de M "unos"; el cabezal de impresión, por ejemplo, avanza o retrocede a razón de un solo cuadro cada vez. Una forma de conseguir el resultado es crear una serie de instrucciones para una máquina de Turing de tres estados. En el primer estado, el cabezal inspecciona la cinta de izquierda a derecha, a razón de un cuadro por vez, hasta encontrar el 1 situado más a la izquierda. La máquina convierte el 1 en un 0, adopta el segundo estado y prosigue su movimiento hacia la derecha. En el segundo estado, cuando el cabezal encuentra un 1 en la casilla bajo inspección, no hace nada –no cambia ni el símbolo de la cinta, ni el estado de la máquina– excepto avanzar una casilla más hacia la derecha. De este modo, el cabezal va dejando atrás los $M - 1$ "unos" que aún quedan en la primera serie. Cuando finalmente aparece un 0, las instrucciones hacen que el cabezal transforme el 0 en un 1 y que la máquina se detenga.

Invitamos ahora al lector para que construya una máquina de Turing capaz de efectuar el producto de dos números dados. Los convenios para la entrada de los números son los mismos que para la máquina sumadora: los dos enteros a multiplicar vienen representados por dos series de "unos" consecutivos, separados por un 0. La salida, una serie de "unos" cuya longitud ha de ser igual al producto de las dos primeras series, puede quedar situada como una tercera serie, a la derecha de las dos primeras. Damos en la figura 5 un diseño capaz de funcionar. Para ser sinceros, es preciso decir que la construcción es delicada y requiere una "contabilidad" algo minuciosa. También parece leal dar una indicación, de la que los aficionados a rompecabezas pueden prescindir saltando el próximo parágrafo.

La multiplicación resulta mucho más



2. CADENAS DE CARACTERES de longitudes finitas cualesquiera forman un conjunto coordinable con el de los enteros positivos, siempre que cada carácter de la cadena sea elegido de entre un conjunto finito. Según definición formulada por Georg Cantor, tal correspondencia biunívoca expresa la igualdad de tamaños de ambos conjuntos, no obstante ser ambos conjuntos infinitos. Cualquier conjunto infinito coordinable con los números enteros positivos se llama "conjunto contable" o "numerable". Dado que todo posible programa de ordenador y toda posible tabla de instrucciones para máquinas de Turing pueden ser codificadas mediante cadenas finitas de símbolos, el emparejamiento muestra que los posibles programas de ordenador, o de máquinas de Turing, forman conjuntos infinitos de tipo numerable.

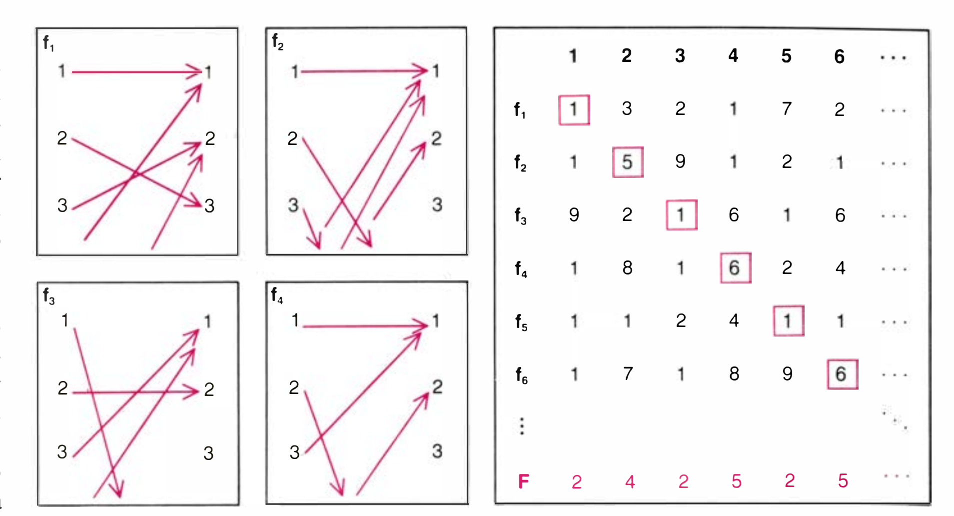
sencilla si primero preparamos una rutina, encargada de copiar cadenas de "unos" inmediatamente a la derecha de un punto dado. Siempre es posible copiar N "unos" utilizando una máquina de N estados; los estados pueden servir de modo efectivo como contadores de los "unos" de la serie. Sin embargo, no siendo deseable que el número de estados crezca ilimitadamente, resulta conveniente dar con un procedimiento para copiar la cadena sin contar sus elementos. Uno de los posibles es hacer que un marcador recorra la serie desde su extremo izquierdo hasta el derecho; tal marcador es un 0 que va ocupando sucesivamente el lugar de cada 1 de la serie. Para cada avance del marcador desde una cuadrícula a la siguiente se instruye al cabezal que inspeccione la serie hasta rebasar su extremo derecho, que pase por alto el cero indicador de espacio en blanco y que cambie después el siguiente 0 que encuentre, convirtiéndolo en un 1. Si el cabezal de impresión va y viene una vez, a modo de lanzadera de un telar por cada avance del marcador, quedará escrita una nueva serie de N "unos" consecutivamente a la derecha de la primera serie [*véase la figura 4*].

Construcción de programas complejos

Con algo de práctica se aprende pronto a construir máquinas de Turing capaces de ejecutar rutinas computacionales sencillas y a combinar entre sí estas máquinas a fin de realizar cálculos más complejos. Por ejemplo, los valores de una expresión polinómica pueden ser evaluados combinando rutinas de adición, copiado y multiplicación. Todavía más versátiles son rutinas breves y elementales dedicadas a manipulación de símbolos, tales como "desplazamiento del cabezal hacia la derecha, hasta que encuentre un 0" o "desplazamiento del marcador de la serie de unos situada más a la izquierda una cuadrícula hacia la derecha". Se saca provecho de variantes de estas rutinas breves tanto en la máquina copiadora como en la multiplicadora.

Si nos sentimos inclinados a construir, pongamos por caso, la máquina de Turing capaz de multiplicar, pronto comenzaremos a apreciar las dificultades que es preciso afrontar al diseñar un programa de utilidad para un ordenador. La mayoría de las máquinas de Turing pequeñas, esto es, que tan sólo admiten un reducido número de estados, son incapaces de realizar tareas útiles, y, muchas veces, ni siquiera tareas razonables. Una buena parte de ellas quedan atrapadas en bucles infinitos, oscilando indefinidamente adelante y atrás de la cinta, sin jamás llegar a detenerse. De entre las máquinas que sí ejecutan tareas razonables debemos elegir una combinación de ellas capaces de colaborar eficientemente. La impresión inicial puede ser que la tarea más nimia puede plantear la dificultad más perversa, y que apenas si hay esperanza de computación en casos tomados de la realidad. Tales dificultades son sin duda enojosas, pero su carácter es meramente técnico; con unas cuantas rutinas bien elegidas la capacidad de resolución de problemas que adquiere la máquina de Turing se multiplica explosivamente, y no es ahora la debilidad de la máquina, sino su potencial, el que nos deja pasmados. Como Turing pudo demostrar, es posible combinar máquinas sencillas y combinar con ellas una máquina capaz de llevar a cabo cualquier tarea que admita descripción secuencial explícita.

El ordenador electrónico, que debe en parte su existencia a las máquinas conceptuales de Turing, es hoy, con toda probabilidad, la demostración más convincente de la potencia computacional de aquellas máquinas. En el curso de su trabajo, Turing señaló que cualquier máquina de Turing M puede quedar codificada en forma de sucesión de unos y ceros impresos en una cinta. La razón fundamental de que tal codificación puede realizarse es que toda máquina de Turing queda unívocamente definida por su tabla de instrucciones, tabla que ha de ser de longitud finita, ya que tanto el número de estados de la máquina como el alfabeto de símbolos a grabar son finitos.



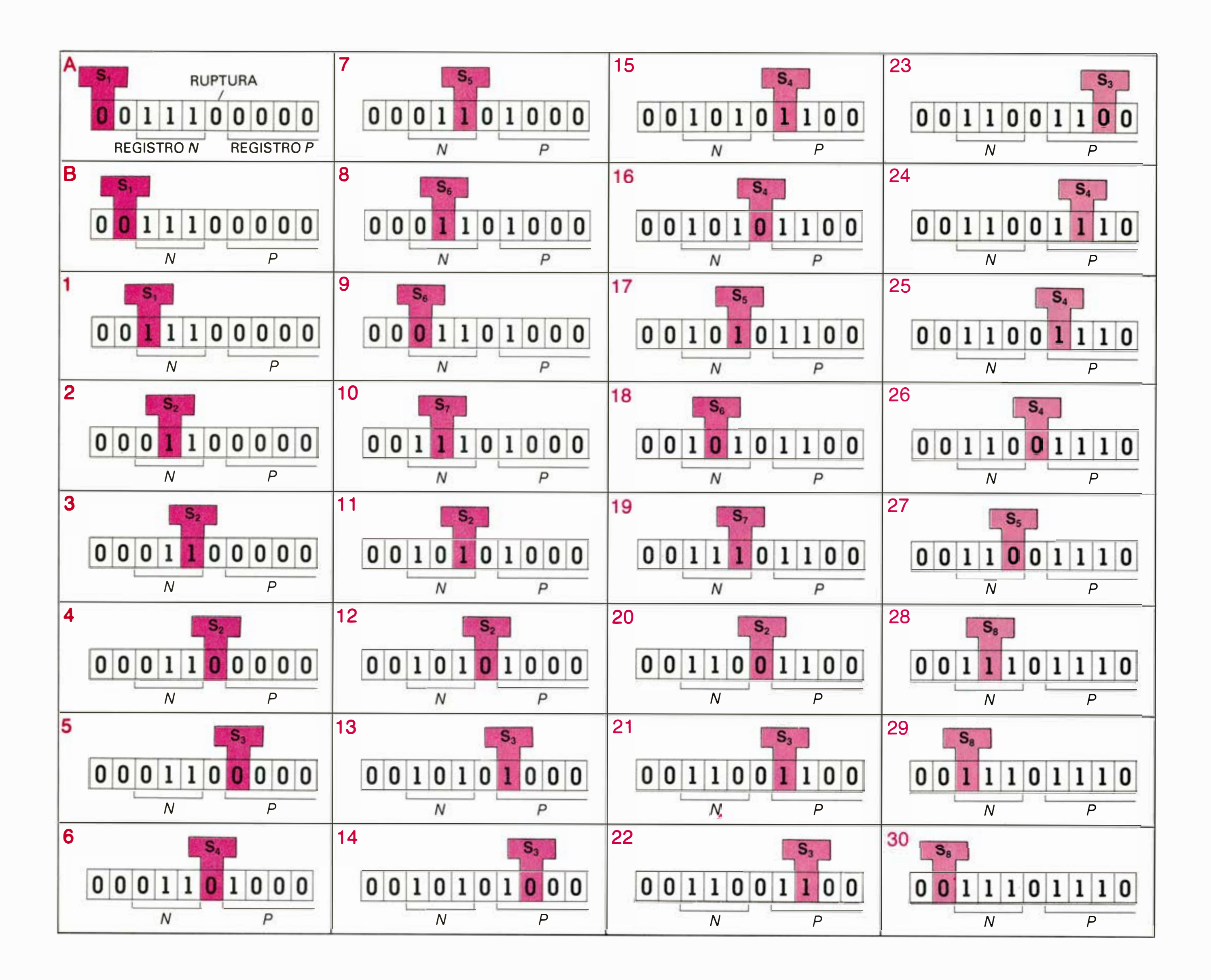
	1	2	3	4	5	6	...
f_1	1	3	2	1	7	2	...
f_2	1	5	9	1	2	1	...
f_3	9	2	1	6	1	6	...
f_4	1	8	1	6	2	4	...
f_5	1	1	2	4	1	1	...
f_6	1	7	1	8	9	6	...
⋮							⋱
F	2	4	2	5	2	5	...

3. EL CONJUNTO DE FUNCIONES MATEMATICAS con valores enteros que pueden definirse sobre el conjunto de los enteros positivos no es numerable; hay demasiadas de tales funciones como para emparejarlas una a una con números naturales. Supongamos que, por el contrario, todas las funciones tales pudieran ser nombradas asociándole un subíndice a cada una, con lo que podríamos enumerarlas todas en un orden fijo. Los valores que cada función asocia a cada número entero podrían ser entonces presentados en una hilera de infinitas cifras. Dado que una función está determinada por su hilera infinita de dígitos, imaginemos una función construida cambiando la primera cifra de la primera hilera, la segunda cifra de la segunda hilera de dígitos, y así sucesivamente. La hilera de dígitos correspondientes a la función así construida (*en color*) se diferencia al menos en una cifra de cualquier otra hilera de la lista original. La hipótesis según la cual todas las funciones pueden ser enumeradas en una lista engendra la contradicción de que siempre puede construirse una función nueva, no incluida en la lista primitiva.

La máquina universal de Turing

Turing demostró que la actuación de una máquina de Turing arbitraria M sobre una sucesión cualquiera X de símbolos de su cinta puede ser simulada por otra máquina, la máquina universal de Turing. Los símbolos que la máquina universal va registrando y reconociendo en su cinta están agrupados en dos secciones principales: a la izquierda está la descripción codificada de la máquina M; a la derecha está la sucesión de símbolos X que M encontraría al ir inspeccionando la cinta que le es propia. La máquina universal se construye entonces de modo que su cabezal vaya y venga entre las dos secciones, izquierda y derecha, de la cinta. Mediante un complejo sistema de marcadores, la máquina universal lleva la cuenta del estado codificado de M que se está consultando. Turing demostró que el efecto de la máquina universal sobre la sucesión de símbolos X es exactamente el que produciría M sobre la misma sucesión.

El éxito de la simulación que la má-



ESTADO	SIMBOLO DE LA CINTA: 0	COMENTARIO	EJEMPLOS	SIMBOLO DE LA CINTA: 1	COMENTARIO	EJEMPLOS
S_1	0,S_1,D	Se comienza con el desplazamiento hacia la derecha, hasta el límite izquierdo del registro *N*.	A, B	0,S_2,D	Marcar con 0 el extremo izquierdo del registro *N*.	1
S_2	0,S_3,D	Desplazamiento hacia la derecha, a través de la ruptura.	4, 12, 20	1,S_2,D	Desplazamiento hacia la derecha, a través del registro *N*, desde el marcador hasta la ruptura.	2, 3, 11
S_3	1,S_4,I	Copiar un 1 en el extremo derecho del registro *P*.	5, 14, 23	1,S_3,D	Desplazamiento hacia la derecha, a través de los "unos" del registro *P*.	13, 21, 22
S_4	0,S_5,I	Desplazamiento hacia la izquierda, a través de la ruptura.	6, 16, 26	1,S_4,I	Desplazamiento hacia la izquierda, a través del registro *P*.	15, 24, 25
S_5	1,S_8,I	El marcador 0 es detectado inmediatamente a la izquierda de la ruptura, lo que indica que al registro *P* se le ha añadido una copia completa del registro *N*.	27	1,S_6,I	Desplazamiento de un paso hacia la izquierda, a partir de la ruptura, entrando en el registro *N*. El estado no puede continuar en S_5. De hacerlo, el encuentro que llegaría a producirse con el 0 haría que la máquina se detuviese prematuramente.	7, 17
S_6	1,S_7,D	Iniciar el desplazamiento del marcador 0 un paso hacia la derecha.	9, 18	1,S_6,I	Proseguir el desplazamiento a través del registro *N*, hasta el marcador 0.	8
S_7		No se encuentra condición; por tanto, puede insertarse una instrucción ficticia cualquiera o no insertarse ninguna.		0,S_2,D	Terminar el desplazamiento del marcador 0 una unidad hacia la derecha. Repetir el ciclo.	10, 19
S_8	0,–,–	La ruptura de copiado está terminada. Detención de la máquina o espera de nuevas instrucciones.	30	1,S_8,I	Desplazamiento hacia la izquierda, hasta el extremo izquierdo del registro *N*.	28, 29

4. MAQUINA COPIADORA, máquina de Turing que se utiliza como componente de otros dispositivos más complejos. Dada una cadena cualquiera de "unos" impresos en la cinta, la máquina escribe a continuación del 0 que señala el fin de la cadena una segunda serie con igual número de "unos".

quina universal hace de la máquina *M* es consecuencia tan sólo de que *M* pueda quedar exhaustivamente descrita por un número finito de símbolos. En principio, no obstante, todo ordenador digital puede ser descrito de igual manera. El ordenador tiene un número grande, aunque finito, de estados internos, y su respuesta a los datos ingresados en él queda totalmente determinada por el conjunto finito de enunciados que componen su programación. Por tanto, la descripción completa de cualquier ordenador digital puede quedar codificada en forma de sucesión de unos y ceros grabados sobre una cinta, y cualesquiera datos de entrada, quedar codificados en la misma cinta, a la derecha de la descripción del ordenador. Consultando alternativamente la descripción del ordenador y la serie de datos de entrada grabados en la cinta, la máquina universal ideada por Turing puede simular, paso por paso, la acción del ordenador sobre los datos de entrada.

Proveyéndolo de memoria suficiente, que haga el papel de cinta con la que manipular los símbolos, cualquier ordenador real puede desempeñar el papel de máquina universal de Turing. Por ejemplo, si un ordenador doméstico fuese programado para funcionar como máquina universal de Turing, y suponiendo que una descripción de un gran ordenador "de cabecera" le fuese codificada al pequeño como datos de entrada, el microordenador simularía la actuación que el ordenador grande pudiera tener sobre cualquier serie de símbolos-datos. En este sentido, todos los ordenadores digitales pueden computar exactamente la misma clase de funciones matemáticas, a saber, todas las que sean computables por algunas máquinas de Turing. La existencia de tan sólo una clase tal respalda con fuerza la definición formal de computabilidad dada por Turing: una función matemática es computable si puede ser computada por alguna máquina de Turing. Turing arguyó convincentemente que su definición equivale a cualquier interpretación razonable de la noción intuitiva de computabilidad. Tal vez convenga mencionar, no obstante, que es absurdo pedir una demostración matemática rigurosa de que una definición formal, como la de Turing, capta perfectamente una cierta noción de carácter intuitivo.

El programa de Hilbert

Para comprender por qué estaba Turing tan empeñado en definir qué es la computabilidad conviene tener alguna noción del desarrollo de la lógica matemática hasta 1936. La matemática rigurosa, en la forma en que hoy es conocida, es cosa relativamente reciente. El primer intento serio de reducción de los enunciados matemáticos a enunciados de la lógica formal fue el iniciado por Gottlob Frege en 1879, con la publicación de su *Begriffsschrift* ("La notación de los conceptos", literalmente). El problema planteado por Hilbert tenía, pues, trascendencia inmediata para los matemáticos, si el esquema de Frege podía llevarse a conclusión. En efecto, si pudiera darse con un método para determinar la veracidad o falsedad de un enunciado cualquiera de la lógica formal, y de ser correctas las ideas de Frege, con tal método se podría determinar también la veracidad o falsedad de cualquier enunciado matemático, no importa cuán complejo. Con un tal método quedarían inmediatamente liquidadas conjeturas como "el último teorema de Fermat" que durante siglos han venido resistiendo todos los embates para demostrarlas o refutarlas. Si el osado reto lanzado por Hilbert tuviera respuesta afirmativa, toda la matemática habría quedado reducida a cálculo mecánico.

Durante los primeros decenios de nuestro siglo se produjeron en lógica dos importantes desarrollos, que desorganizaron y echaron por tierra gran parte del programa de Hilbert. En 1901, Bertrand Russell descubrió en la teoría elemental de conjuntos una paradoja irrefutable, teoría elemental que desempeñaba un papel fundamental en el programa de reducción de la matemática a la lógica propuesto por Frege. Russell comunicó a Frege su descubrimiento justamente cuando el segundo volumen de la última gran obra de Frege, *Grundgesetze der Arithmetik* ("Las leyes fundamentales de la aritmética"), estaba a punto de ser publicado. Frege cerró el volumen con una desalentada nota: "Difícilmente puede un científico encontrarse nada más indeseable que ver hundirse los cimientos justamente cuando se está concluyendo el trabajo. Esta ha sido la situación en que me he encontrado a causa de una carta del Sr. Bertrand Russell, recibida cuando la obra estaba a punto de salir de imprenta". A pesar de los fallos, que Frege no pudo corregir en su obra, Russell y Alfred North Whitehead pudieron más tarde salvar el programa de Frege y eludir la paradoja en la teoría de conjuntos.

Un segundo y fundamental descubrimiento en lógica fue el realizado por Kurt Gödel, a la sazón en la Universidad de Viena. Implícitamente, el programa de Hilbert presuponía que en la lógica formal debiera existir un procedimiento para distinguir los enunciados verdaderos de los falsos; el problema consistía en dar con el método. Gödel demostró que tal presunción es injustificada. En 1931 pudo probar que cualquier sistema consistente de la lógica formal que fuera lo bastante potente como para formular en él enunciados acerca de la teoría de números (la aritmética) ha de contener enunciados verdaderos que no pueden ser demostrados. Dado que sistemas axiomáticos consistentes, tales como el ideado por Russell y Whitehead, no pueden contener todos los enunciados verdaderos de la teoría que pretenden formalizar, se dice que tales sistemas son incompletos.

La lógica de la computabilidad

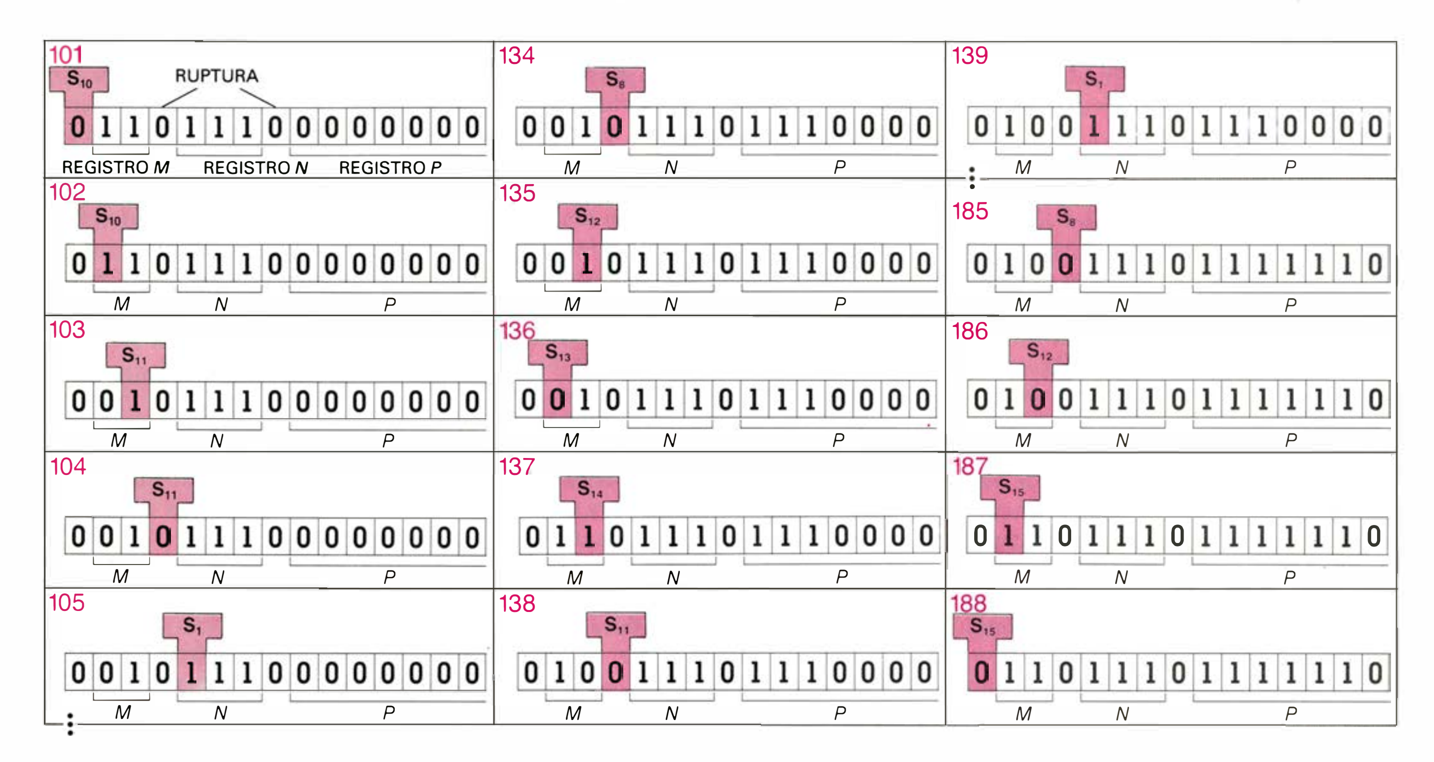
El trabajo de Gödel puso verdaderamente término al programa de Hilbert. No puede existir un método que permita decidir si un enunciado arbitrariamente dado es verdadero o falso. De haberlo, el método constituiría una demostración de todos los enunciados verdaderos, y Gödel había demostrado que en el seno de un sistema axiomático consistente, de potencia suficiente para abarcar la aritmética ordinaria, una tal demostración es imposible. La atención de los lógicos se desplazó desde la noción de verdad a la de demostrabilidad. Y en este contexto quedaba por elucidar una cuestión, análoga sencilla de la de Hilbert; a saber: ¿Existe un único método mediante el cual todos los enunciados matemáticos demostrables puedan ser demostrados a partir de un sistema de axiomas lógicos?

El más distinguido investigador de la lógica de demostrabilidad, de los años inmediatamente posteriores a la demostración de Gödel, fue Alonzo Church, de la Universidad de Princeton. Church y dos de sus discípulos, Stephen C. Kleene y J. Barkley Rosser, pusieron a punto un lenguaje formal coherente, bautizado por ellos "lambda-cálculo". Este lenguaje es útil para razonar sobre funciones matemáticas, tales como la raíz cuadrada, los logaritmos y cualesquiera otras funciones más complicadas que pudieran definirse. (Church eligió la letra griega "lambda", equivalente a la *L* latina, para sugerir que su sistema formal es un lenguaje). El moderno lenguaje de programación llamado Lisp (de *Lis*t

*p*rocessing, procesamiento de listas) está inspirado en el lambda-cálculo. Kleene demostró que amplias clases de funciones matemáticas, incluidas todas las utilizadas por Gödel en su demostración, podrían ser expresadas en el lambda-cálculo.

El siguiente paso importante en esta misma línea de pensamiento fue dado por Church. Aducía que si una función matemática es computable –queriendo con ello decir que puede calcularse su valor para todo número perteneciente a su dominio de definición– entonces tal función es definible en el lambda-cálculo. El trabajo de Church mostraba que de haber cosa tal como una función matemática expresable en el lambda-cálculo, pero no computable, no habría método alguno para determinar si un enunciado matemático dado es o no demostrable, y no digamos verdadero. La última de las hipótesis aún supervivientes del programa de Hilbert quedaría al fin infirmada: en abril de 1936, Church publicó una fórmula lógica que no es computable en su sistema.

Turing, trabajando independientemente, y sin relación con Church, había comprendido también que existía una relación técnica entre el problema de Hilbert y la noción de función computable. Empero, al atacar el problema



ESTADO	SIMBOLO DE LA CINTA: 0	COMENTARIO	EJEMPLOS	SIMBOLO DE LA CINTA: 1	COMENTARIO	EJEMPLOS
S_{10}	0,S_{10},D	Se comienza con desplazamiento hacia la derecha, hasta el límite izquierdo del registro *M*.	101	0,S_{11},D	Marcar con un 0 el extremo izquierdo del registro *M*.	102
S_{11}	0,S_{1},D	Desplazamiento hacia la derecha, a través de la ruptura que separa los registros *M* y *N*. Comienza la rutina de copiado (*véase la ilustración precedente*).	104, 138	1,S_{11},D	Desplazamiento hacia la derecha, a través del registro *M*, desde el marcador hasta la ruptura.	103
S_{8}	0,S_{12},I	Desplazamiento hacia la izquierda, a través de la ruptura entre *M* y *N*. Fin de la rutina de copiado.	134, 185	1,S_{8},I	Desplazamiento hacia la izquierda, hasta el extremo izquierdo del registro *N*.	No figuran
S_{12}	1,S_{15},I	El marcador 0 es detectado inmediatamente a la izquierda de la ruptura entre los registros *M* y *N*, lo que indica que el producto de los números almacenados en los registros *M* y *N* ha sido almacenado en el registro *P*.	186	1,S_{13},I	Desplazamiento izquierdo, desde la ruptura entre los registros *M* y *N*, entrando un paso en el registro *M*. El estado no puede seguir siendo S_{12}; de serlo, el encuentro con el marcador 0 haría detenerse prematuramente la máquina de Turing.	135
S_{13}	1,S_{14},D	Comienza el avance del marcador 0 situado en el registro *M*, un paso hacia la derecha.	136	1,S_{13},D	Prosigue el desplazamiento hacia la izquierda, a través del registro *M* hasta el marcador 0.	No figuran
S_{14}		No se encuentra condición. Puede, por tanto, insertarse una instrucción ficticia cualquiera, o no insertarse ninguna.		0,S_{11},D	Terminar el avance del marcador 0 del registro *M*, una unidad hacia la derecha; repetir el ciclo.	137
S_{15}	0,–,–	La rutina de multiplicación está terminada. Detención de la máquina, o espera de nuevas instrucciones.	188	1,S_{15},I	Desplazamiento hacia la izquierda, hasta el extremo izquierdo del registro *M*.	187

5. PODEMOS EFECTUAR LA MULTIPLICACION mediante una máquina de Turing integrando en ella otra máquina de Turing capaz de copiar cadenas de "unos". En este ejemplo se requieren 88 ciclos para hallar el producto de 2 por 3. Los saltos de numeración corresponden a ciclos de máquina que trasladan el control a estados definidos para la máquina copiadora (*véase la figura precedente*). En el último ciclo se presenta al producto de 2 por 3 en forma de cadena de seis "unos" impresos en una sección de la cinta llamada registro *P*, que se halla situado inmediatamente detrás del 0 que separa los dos factores.

UN SIMBOLO

1 2 3 9 π

DOS SIMBOLOS

11 99 9! $(=1\times2\times3\times4\times5\times6\times7\times8\times9=362{,}880)$ 9^9 $(=387{,}420{,}489)$

TRES SIMBOLOS

UNO 9+9 $3\uparrow4$ $(=3\times[3\times(3\times3)]=3^4=81)$ 9^{9^9} $(=9^{387,420,489}\cong10^{369,700,000})$

CUATRO SIMBOLOS

TRES 10^{99} $3\uparrow\uparrow4$ $(=3\uparrow[3\uparrow(3\uparrow3)]=3^{3^{3^3}}=3^{3^{27}}=3^{7,625,597,484,987}$
$\cong10^{3,638,000,000,000})$

CINCO SIMBOLOS

CINCO $3\uparrow\uparrow\uparrow4$ $(=3\uparrow\uparrow[3\uparrow\uparrow(3\uparrow\uparrow3)]=3\uparrow\uparrow[3\uparrow\uparrow3^{3^3}]=3\uparrow\uparrow3^{3^{\cdot^{\cdot^{3}}}}\Big\}3^{3^3}$ NIVELES

$= 3^{3^{\cdot^{\cdot^{3}}}}\Big\}3^{3^{\cdot^{\cdot^{3}}}}$ NIVELES $\Big\}3^{3^3}$ NIVELES)

⋮

53 SIMBOLOS

1#MAS#QUE#EL#MAXIMO#NUMERO#EXPRESABLE#CON#53#SIMBOLOS

6. PARADOJA DE RICHARD, así llamada en honor del matemático francés Jules Richard. Se suscita al suponer que los enteros positivos pueden ser ordenados, o numerados en lista, de acuerdo con el número de símbolos necesarios para especificarlos. Es posible designar enteros aislados de considerable tamaño con ayuda de símbolos especiales, como la notación sagital ideada por Donald E. Knuth, de la Universidad de Stanford. Empero, de acuerdo con su propia descripción, el número supuestamente especificado arriba mediante 53 símbolos es mayor que cualquier otro que pueda ser especificado mediante 53 símbolos, lo que no deja de constituir una paradoja. Se presentan paradojas similares cuando se intenta hallar una máquina de Turing determinada en una lista de máquinas de Turing que se encuentran dispuestas según la longitud de la cadena de símbolos necesarios para codificar cada una de dichas máquinas.

procedió de forma mucho más directa y concreta que Church. Lo que se requería era un modelo a la vez preciso y sencillo del proceso de computación; las máquinas de Turing fueron diseñadas para satisfacer tal necesidad. Una vez especificadas cuáles habrían de ser sus propiedades, Turing estableció brillantemente la conexión entre la idea de función computable y los resultados de trabajos matemáticos realizados unos 50 años antes por el alemán Georg Cantor. Cantor había razonado que si bien no existe un número entero que sea el máximo, cualquier conjunto infinito de objetos que pudieran ser contados, esto es, biunívocamente asociados con los números enteros positivos, es un conjunto del mismo tamaño que el conjunto de los números enteros. Dado que cualquiera de las máquinas de Turing puede ser expresada por una cadena de caracteres de longitud finita, todas las posibles máquinas de Turing, y con ellas todas las posibles funciones computables, pueden ser inventariadas en orden numérico o lexicográfico; pueden por tanto ser puestas en correspondencia biunívoca con los números enteros [*véase la figura 2*]. Como es obvio, no existe una cota superior que limite el tamaño de las máquinas de Turing, ni por consiguiente tampoco hay límite para los números de posibles máquinas de Turing distintas. No obstante, el análisis de Cantor muestra que el conjunto de todas las posibles funciones computables es del mismo tamaño que el conjunto de todos los números enteros; ambos conjuntos son llamados contables, o numerables.

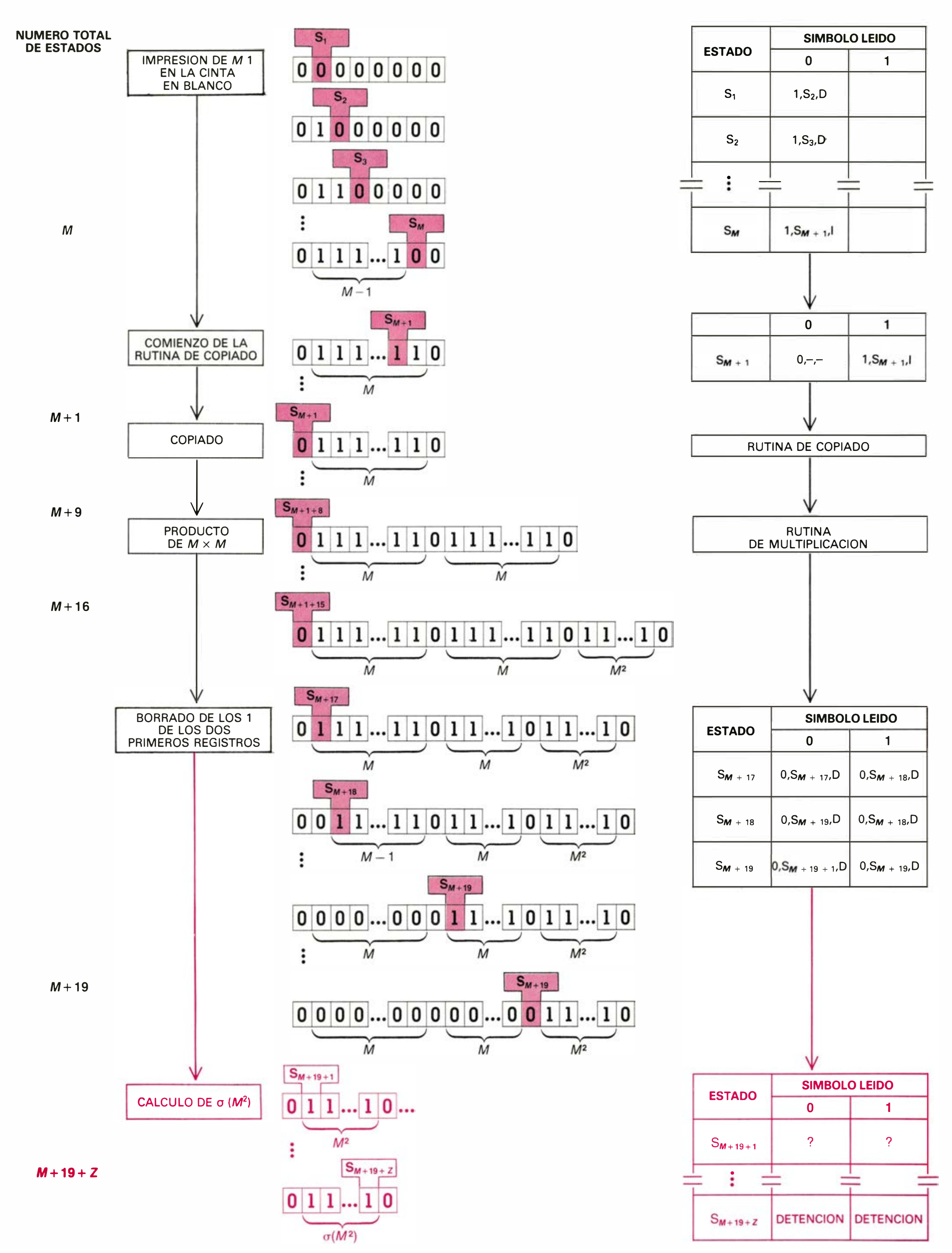
Cantor había demostrado también que hay infinitos conjuntos que no son numerables; son conjuntos mayores que el conjunto de los números enteros, en el sentido de que no pueden ser biunívocamente asociados con éstos. Entre los ejemplos de tales conjuntos no numerables tenemos el conjunto de todas las funciones dependientes de números enteros, con valores también enteros. Un análisis cuidadoso demuestra que ha de haber más de tales funciones que números enteros. La consecuencia es que será imposible que existan suficientes programas como para computar toda posible función.

El problema de la detención

¿Cuáles son las funciones no computables? Desdichadamente, las demostraciones de Church y de Turing no proporcionan directamente ejemplos de funciones no computables. No obstante, durante los veinte últimos años, se ha sacado partido de las máquinas de Turing para construir varias de tales funciones. Uno de los primeros ejemplos de función no computable fue ideado por Tibor Rado, de la Universidad estatal de Ohio, en 1962. Fijémonos en todas las máquinas de Turing que tienen cierto número fijo N de estados. Supongamos que todas las máquinas se hacen comenzar a funcionar sobre una cinta en blanco, o más precisamente, sobre una cinta cuyos cuadros están todos ocupados por ceros. Imagi-

NUMERO DE ESTADOS	NUMERO MAXIMO DE "UNOS" IMPRESOS	COTA INFERIOR PARA EL VALOR DE σ
3	$\sigma(3)$	6
4	$\sigma(4)$	12
5	$\sigma(5)$	17
6	$\sigma(6)$	35
7	$\sigma(7)$	22,961
8	$\sigma(8)$	$3^{92} \cong 7.9 \times 10^{43}$
9	$\sigma(9)$	$3^{92}+1$
10	$\sigma(10)$	$a^{a^{\cdot^{\cdot^{a}}}}\Big\} a^{a^{\cdot^{\cdot^{a}}}}\Big\} \cdots \Big\} a^a$

7. EL PROBLEMA DEL "CASTOR AFANOSO" consiste en hallar el número máximo de "unos" que puede llegar a imprimir una máquina de Turing de N estados que comienza a funcionar con una cinta inicialmente ocupada sólo por "ceros" y que ha de terminar por detenerse. Tal número, que depende de N, constituye el valor de una función designada por $\sigma(N)$. En 1962, Tibor Rado, de la Universidad estatal de Ohio, demostró que tal función crecía con excesiva rapidez para poder ser computable. Vemos en la tabla cotas inferiores de la función, estimadas para valores pequeños de N. La cota inferior correspondiente a $\sigma(10)$ puede ser expresada mediante un número a cuyo valor es, aproximadamente, $\sqrt{8}$; $\sigma(10)$ es una pila de a-es exponencialmente acumuladas, donde el número de a-es de cada pila viene expresado por otro rimero de a-es. El proceso de definición de la altura de una pila mediante otra se reitera 10 veces.



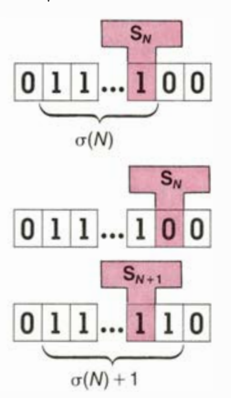
8. LA DEMOSTRACION DE NO COMPUTABILIDAD de σ(N) comienza por calcular una estimación del número de estados necesarios para generar una cadena de M^2 signos 1 en una cinta inicialmente ocupada por "ceros". Las máquinas de Turing para copiar y multiplicar han sido combinadas aquí, formando una máquina de $M + 16$ estados. Otros tres estados más se encargan de borrar las dos cadenas de "unos" situadas a la izquierda, dejando una cadena de M^2 cifras 1. Se supone que para toda cadena de N cifras 1, hay una máquina de Z estados capaz de generar σ(N) "unos" (*texto en color*).

nemos que de este conjunto hayamos podido excluir todas las máquinas de Turing que no llegan jamás a detenerse. De entre las restantes, seleccionemos la máquina o grupo de máquinas que hayan impreso sobre la cinta en blanco el máximo número de "unos" en sucesión antes de detenerse. Tal número de "unos", que está definido para cada valor de N, es el valor de la función de Rado, y es habitualmente designado por $\sigma(N)$. La demostración detallada de que $\sigma(N)$ no es computable consiste en suponer que sí puede ser computada y en deducir una contradicción. El razonamiento es bastante directo, pero los detalles técnicos son intrincados [*véanse las figuras 8 y 9*].

Pudiera pensarse que el razonamiento es deficiente, porque presume que las máquinas de Turing de N estados que no llegarían a detenerse pueden ser eliminadas de antemano. Tal objeción es verdaderamente seria. Fijémonos, así pues, en cómo podríamos intentar el cálculo de $\sigma(N)$ por la fuerza bruta. Damos primero la lista completa de todas las máquinas de Turing de N estados, ordenadas según cierto criterio. Simulemos cada una de ellas en una máquina universal de Turing; y seleccionemos la máquina o grupo de máquinas que más "unos" hayan impreso. Aunque este método de computación parezca eliminar la objeción, la dificultad planteada por las máquinas de Turing que no llegan a detenerse reaparece, esta vez en forma de no-tratable. Algunas de las máquinas de N estados que llegan a detenerse pueden quedar eliminadas mediante algoritmos sencillos, pero hay otras para las que no puede tomarse tal decisión. Si no podemos determinar si una máquina concreta se va o no a parar, no podremos eliminarla de la lista de máquinas de N estados y la simulación habrá de continuar. Y dado que muy bien puede suceder que la máquina no llegue a detenerse jamás, no existe garantía de que el cálculo de $\sigma(N)$ pueda llevarse a término.

Si bien la máquina de Turing produjo su impacto primero en la lógica matemática, ha desempeñado un papel dominante en informática desde hace unos 20 años. En 1965, Juris Hartmanis y Richard E. Stearns, a la sazón en los General Electric Research Laboratories, en Schenectady, Nueva York, demostraron que la máquina de Turing podía ayudar a establecer acotaciones muy estrictas de complejidad de computación. Investigadores subsiguientes comenzaron a clasificar los problemas según la forma de dependencia del tiempo de funcionamiento, o lo que

"Observemos ante todo que σ es uniformemente creciente: es decir, si $X > Y$, $\sigma(X) > \sigma(Y)$. Si una máquina de Turing de N estados imprime un número dado de "unos", siempre podrá construirse una máquina de $N + 1$ estados que añada un 1 a la anterior cadena de 1s, y después se detenga."



ESTADO	SIMBOLO LEIDO	
	0	1
S_N	$1, S_{N+1}, I$	$1, S_N, D$
S_{N+1}	DETENCION	DETENCION

Supongamos que exista una máquina de Turing, de Z estados, capaz de computar $\sigma(M^2)$.

Por la definición de σ y el cálculo mostrado en la ilustración de la página adyacente,

$\sigma(M + 19 + Z) \geqslant \sigma(M^2)$. (1).

Sin embargo, si M es suficientemente grande, el enunciado (1) contradice que σ sea uniformemente creciente.

Para demostrarlo, sea $M = Z + 20$, o sea, $Z = M - 20$

Entonces, $M + 19 + Z = M + 19 + M - 20 = 2M - 1$.

Como es evidente, $2M - 1 = M^2 - (M - 1)^2$, y así $M + 19 + Z = M^2 - (M - 1)^2$.

Dado que $(M - 1)^2 > 1$, $M^2 - 1 > M^2 - (M - 1)^2$.

Así pues, por ser σ uniformemente creciente,

$\sigma(M^2 - 1) > \sigma(M^2 - [M - 1]^2) = \sigma(M + 19 + Z) \geqslant \sigma(M^2)$.

Sin embargo, como $\sigma(M^2 - 1) > \sigma(M^2)$, σ no puede ser uniformemente creciente, y por tanto, la suposición de que existe una máquina de Turing de Z estados que computa $\sigma(M^2)$ implica una contradicción.

9. PASOS FINALES de la demostración de no computabilidad de $\sigma(N)$: extraen una contradicción de la hipótesis de que hay una máquina de Z estados tal que con una cadena de M^2 cifras 1 computa $\sigma(M^2)$.

viene a ser equivalente, del número de pasos de cómputo, respecto del tamaño del problema. Por ejemplo, supongamos que cierto número de puntos N estén interconectados mediante líneas, formando un grafo compuesto por vértices y lados. El problema consiste en colorear los vértices de modo que ningún par de vértices conectados por una línea sean de un mismo color. Supongamos, además, que el más rápido de los métodos conocidos para resolver el problema exija tiempos que varíen según cierta potencia de N, N^2, pongamos por ejemplo. Se dice entonces que el problema pertenece a la clase de problemas resolubles "en tiempo polinómico", clase denotada P. La clase P ha ido adquiriendo cada vez mayor importancia; hoy son muchos los especialistas en ciencias de cómputo que consideran no-tratables los problemas no pertenecientes a la clase P.

Teoría moderna de la complejidad

Observemos que, para que un problema sea asignado a la clase P, ningún caso particular del problema puede exigir para su resolución un tiempo mayor que el polinómico. Dicho con otras palabras, el método de solución es determinístico en el sentido de que garantiza la existencia de una solución, y ello en un tiempo inferior al de una cierta potencia fija del tamaño N del problema.

Podemos definir también máquinas de Turing no determinísticas. A tales máquinas se les permite resolver el problema por tanteo: se conjetura una posible solución y se verifica si la conjetura ha sido o no atinada. Por ejemplo, para determinar si un número es compuesto, la máquina no determinística tantea con un supuesto divisor, y si la división resulta exacta, da por comprobado que el número es compuesto. Una máquina determinística, por el contrario, habría de proceder a buscar sistemáticamente un divisor.

El tiempo requerido para resolver un problema con una máquina no determinística viene medido por la longitud del cálculo mínimo; por tanto, la máquina no determinística parecería tener enorme ventaja sobre la determinística. La experiencia ordinaria sugiere que es más fácil comprobar si algo es solución que empezar por encontrarla. Sin embargo, nadie ha podido demostrar que los problemas resolubles en tiempo polinómico con máquinas no determinísti-

cas –que forman la clase de problemas *NP*– sean intrínsecamente de mayor dificultad que los problemas de la clase *P*. La cuestión de si la clase *P* es o no distinta de la clase *NP*, cuestión conocida por "problema *P-NP*", se ha convertido en una de las principales cuestiones no resueltas de las matemáticas de hoy.

En 1970, Stephen A. Cook, de la Universidad de Toronto, dio un importante paso en relación con el problema *P-NP*. Cook estaba investigando el problema de determinación de condiciones bajo las que una proposición lógica compuesta es verdadera. La proposición compuesta que se forma cuando dos proposiciones simples se conectan por la disyuntiva "o" es verdadera cuando una de las proposiciones, o ambas, son verdaderas. Es en general muy difícil determinar el campo de condiciones de veracidad que deben cumplir las proposiciones simples para satisfacer una proposición compuesta o, con otras palabras, los valores de veracidad o falsedad que deben asignarse a las proposiciones simples para que sea verdadera la proposición compuesta. Cook consiguió demostrar que este problema, llamado problema de satisfactibilidad, es tan difícil como cualquier otro problema de la clase *NP*. La condición necesaria y suficiente para que exista un algoritmo eficiente capaz de resolver el problema de satisfactibilidad es que exista un algoritmo tal para cada uno de los problemas de la clase *NP*. Todo problema que posea esta misma propiedad con relación a una clase dada de problemas se dice ser completo para tal clase.

Transcurrió un año hasta que la mayoría de los investigadores empezaran a captar la importancia del resultado de Cook. En 1971, Richard M. Karp, de la Universidad de California en Berkeley, comenzó a preguntar qué otros problemas naturales podrían desempeñar el mismo papel que desempeña el problema de satisfactibilidad con respecto a la clase *NP*. Karp descubrió que muchos problemas importantes de investigación operativa, incluido el de colorear un gráfico con tres colores, son también de la misma dificultad que cualquiera de los problemas que pueden ser asignados a la clase *NP*; con otras palabras, son *NP*-completos. Puede demostrarse directamente, trasladando un problema al dominio del otro, que el problema de coloreado del grafo y el problema de satisfactibilidad son equivalentes [*véase la figura 10*].

Se ha demostrado por métodos similares que varios centenares de problemas, que antes se tuvieron por distintos, son en realidad variantes unos de otros, cuyas diferencias son puramente notacionales. Todos estos problemas son equivalentes al de satisfactibilidad, y son todos, por tanto, *NP*-completos. Se han descubierto otras varias colecciones de tales problemas completos, tanto para la clase *P* cómo para las clases de problemas no tratables, en las cuales el número de pasos que para su resolución requiere una máquina de Turing crece exponencialmente con el tamaño del problema. Empero, parece todavía prematuro intentar el asalto directo del problema *P-NP*. Su dificultad puede, en parte, apreciarse a partir de desarrollos recientes en teoría de computación.

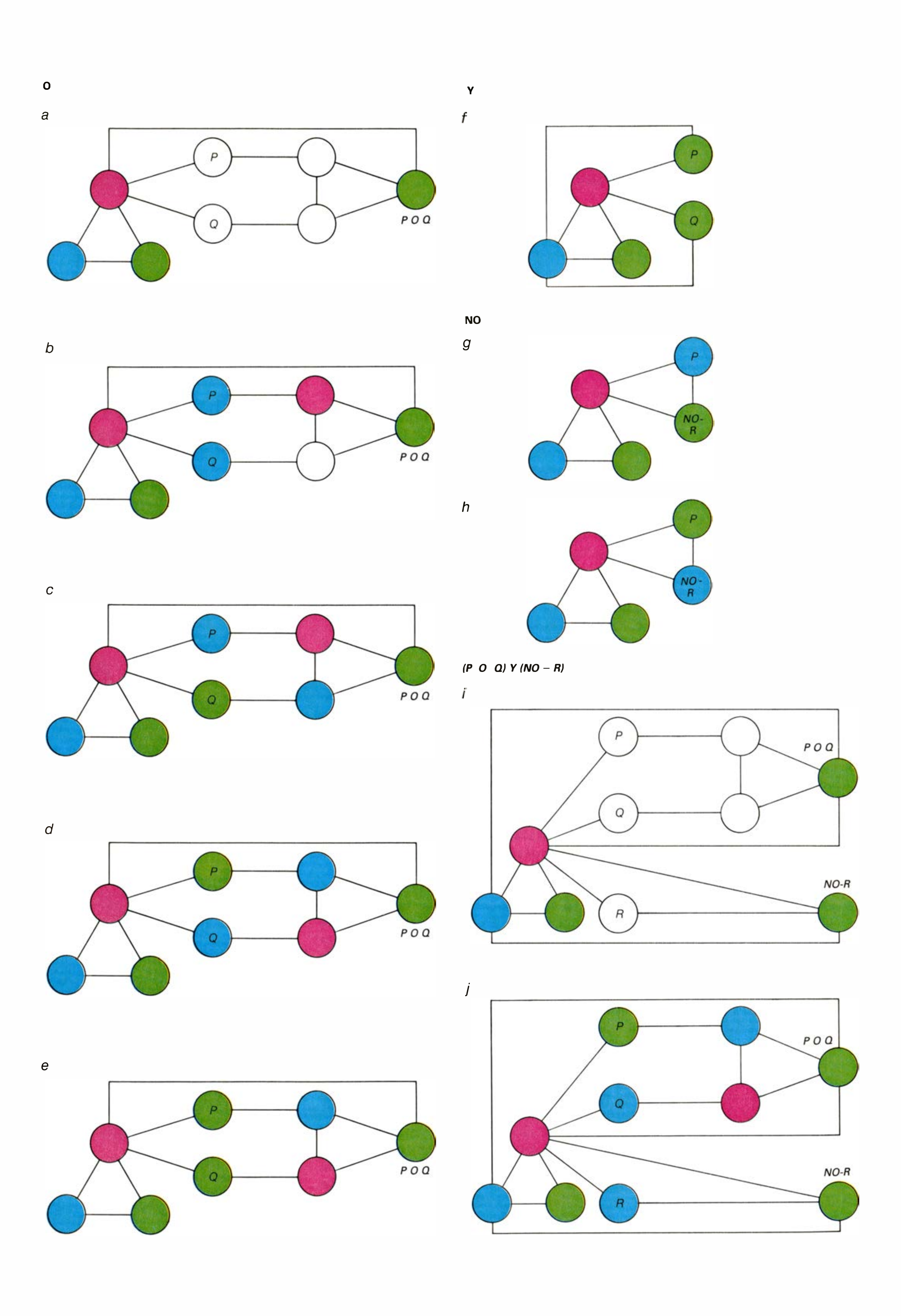
Computabilidad relativa

La idea de que una función pueda ser computada por la máquina de Turing es susceptible de generalización, haciendo que su computabilidad dependa de cadenas de símbolos que la máquina pueda encontrar en su cinta. Si una cadena detectada en la cinta pertenece a una colección *A* de cadenas previamente especificadas, la máquina de Turing puede ser instruida para que avance hasta un estado especial que proceda a computar la función de que se trate. Si la cadena no pertence al conjunto *A*, la máquina dictamina que tal función no es computable. En el primer caso, se dice que la función es computable relativamente al conjunto *A*. Si el conjunto *A* estuviera compuesto por todas las cadenas que codifican máquinas de Turing que llegan a detenerse cuando les proporcionamos como entrada una cinta en blanco, la función de Rado, $\sigma(N)$, sería computable respecto de *A*.

En 1974, Theodore P. Baker, de la Universidad de Cornell, John Gill, de la Universidad de Stanford, y Robert M. Solovay, de la de Berkeley, plantearon si podría demostrarse la inequivalencia de las clases *P* y *NP* para la computación relativa. En el transcurso de su trabajo hicieron un descubrimiento desconcertante: consiguieron especificar dos conjuntos, *A* y *B*, para lo cuales las relaciones entre las clases *P* y *NP* eran contradictorias. Dicho de otra forma, para cómputos relativos al conjunto *A*, las clases *P* y *NP* eran equivalentes, mientras que relativamente al conjunto *B, P* y *NP* eran clases no equivalentes. Se descubrió además que para cualquier sistema formal existían cálculos relativos para los cuales tanto podía suponerse que *P* y *NP* eran equivalentes como que no lo eran, sin detrimento de la coherencia del sistema. Desde entonces, muchos otros investigadores han descubierto numerosos problemas que pueden ser relativizados de modo tal que cada posible resultado sea verdadero.

Esta paradójica situación es obviamente incómoda. Todavía no ha podido resolverse ninguno de los problemas relativizados de dos modos contradictorios, hecho que suele tomarse como prueba de que las soluciones a tales problemas se encuentran más allá del alcance actual de las matemáticas. Empero, debemos recordar que la mera formulación de estos problemas, en apariencia no tratables, ha sido posibilitada por la sencilla solución que se le dio a un problema impenetrable para una generación anterior. Tal vez, visto en retrospectiva, el próximo paso importante llegue a parecernos tan sencillo como las máquinas creadas por la vivaz imaginación de A. M. Turing.

10. SIGNIFICA LA EQUIVALENCIA DE DOS PROBLEMAS que la solución de uno cualquiera de ellos proporciona inmediatamente la solución del otro. Aquí hemos transformado el problema de la determinación de las condiciones en las que es verdadera una proposición lógica o fórmula booleana compleja en otro problema, el de coloreado de los vértices de un grafo con tres colores, de modo tal que ningún par de vértices conectados por una línea sea del mismo color. La veracidad o falsedad de una fórmula booleana cualquiera es función de la veracidad o falsedad de las proposiciones simples, o atómicas, que la componen, y de los modos en que los valores de veracidad están enlazados por las conectivas "o", "y" y "no". A cada posible fórmula booleana se le asocia un gráfico simple, compuesto por vértices y líneas, el cual puede ser coloreado con tres colores si y solamente si pueden asignarse a las proposiciones atómicas valores de verdad de modo tal que la fórmula compuesta sea verdadera. Fijémonos, por ejemplo, en la fórmula compuesta *P o Q*, construida a partir de las proposiciones atómicas *P* y *Q*. La fórmula *P o Q* es verdadera solamente si es verdadera la proposición *P*, o lo es la *Q*, o lo son ambas. Estas condiciones se reflejan en el coloreado del grafo del ángulo superior izquierdo (*a*) de la ilustración, en el cual el verde denota verdad y el azul, falso. El coloreado puede terminarse de modo tal que el vértice en posición extrema derecha sea verde (verdadero) solamente si uno o ambos de los vértices rotulados están coloreados de verde (*b - e*). Análogamente, la fórmula compuesta *P* y *Q* es verdadera solamente si las proposiciones atómicas, *P* y *Q*, son ambas verdaderas, situación que se refleja en un grafo en el cual el vértice que representa a *P* y el vértice que representa a *Q* están ambos coloreados de verde (*f*). Finalmente, la fórmula *no-P* es verdadera solamente si el vértice rotulado *P* está coloreado de azul (*g*); la fórmula *P* es verdadera solamente si el vértice rotulado *no-P* está coloreado de azul (*h*). En el ángulo inferior derecho está el grafo correspondiente a una fórmula booleana más compleja, (*P o Q*) *y* (*no-R*); el coloreado superior (*i*) representa las restricciones impuestas por las conectivas "o", "y" y "no". En la parte baja (*j*) vemos una de las cuatro formas de colorear el grafo, correspondiente a una forma de satisfacer la fórmula booleana.


O
a
P
Q
P O Q
b
P
Q
P O Q
c
P
Q
P O Q
d
P
Q
P O Q
e
P
Q
P O Q
Y
f
P
Q
NO
g
P
NO-R
h
P
NO-R
(P O Q) Y (NO – R)
i
P
Q
P O Q
R
NO-R
j
P
Q
P O Q
R
NO-R

Zonas oceánicas de fractura

Al desencajar los bordes de las placas que constituyen el fondo oceánico complican el modelo concebido por la teoría de la tectónica de placas. La anchura de algunas de ellas alcanza la de una cuenca oceánica

Enrico Bonatti y Kathleen Crane

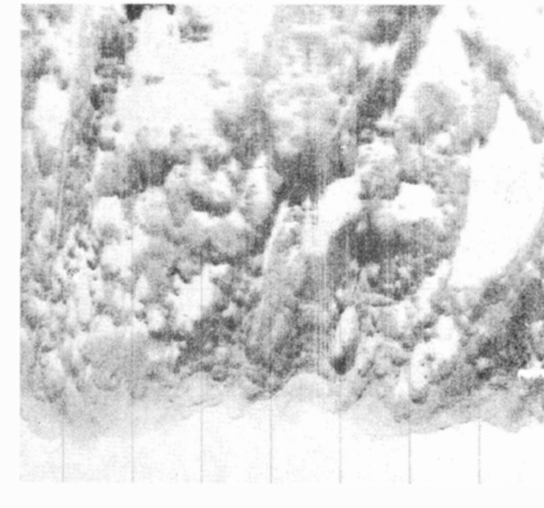
La constatación de que la topografía más accidentada de la superficie del planeta se halla en el fondo de los océanos se remonta a poco más de dos décadas. Primero se descubrió el sistema de dorsales mesoceánicas. Se trata de una cordillera submarina que atraviesa las principales cuencas oceánicas a lo largo de unos 60.000 kilómetros, con lo que constituye, con gran diferencia, la más larga cordillera del globo. Después, la compilación de datos topográficos de las grandes cuencas oceánicas reveló una complicación en su geometría: en muchos lugares, el eje de la dorsal mesoceánica está desencajado lateralmente, de pocos a varios centenares de kilómetros. Los desencajamientos menudean a lo largo de la dorsal mesoatlántica: resiguiendo la cresta de la dorsal se encuentran dislocaciones a intervalos de 50 a 100 kilómetros. La mayoría son cortas: en menos de 30 kilómetros de recorrido se encuentra por dónde continúa la dorsal. Otras son notablemente más largas. Los desencajamientos más largos se caracterizan por la presencia de un surco profundo que une los extremos de dos segmentos de la dorsal. El surco está flanqueado por cumbres, más o menos paralelas al mismo. Se las llama dorsales transversales, o de falla transformante. Destaquemos que el surco y las dorsales de sus flancos suelen poder reseguirse mucho más allá de los segmentos de eje de dorsal que unen. Así pues, el surco y las dorsales transversales constituyen interrupciones generalizadas del fondo oceánico, conocidas hoy por zonas oceánicas de fractura.

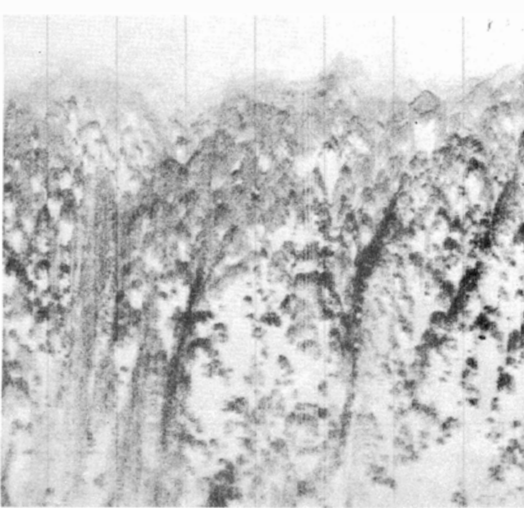
Situémonos en el Atlántico ecuatorial. Un juego de zonas de fractura muy próximas entre sí corta la dorsal mesoatlántica. La mayor de ellas, la zona de fractura de Romanche, disloca el eje de la dorsal casi 1000 kilómetros. Las partes más profundas del fondo de la fosa de Romanche se hallan a más de siete kilómetros bajo el nivel del mar; las partes más elevadas de las dorsales que flanquean la fosa quedan a menos de un kilómetro del nivel del mar. El relieve vertical, por tanto, supera los seis kilómetros. El Gran Cañón del Colorado apenas alcanza una cuarta parte de ese desnivel. Junto a la zona de fractura de Romanche existen otras zonas parecidas, igualmente impresionantes; el conjunto forma una serie de surcos y dorsales transversales de varios centenares de kilómetros de anchura de norte a sur. El terreno resultante, la zona de megacizalla ecuatorial, apenas tiene parangón en tamaño y carácter accidentado en otra parte del planeta.

¿Qué procesos geológicos crearon las zonas de fractura del Atlántico ecuatorial, así como otras zonas de fractura parecidas descubiertas por doquier en las cuencas oceánicas? La respuesta exige el examen de los acontecimientos que producen una cuenca oceánica, del que se concluye que las zonas de fractura son lugares de una actividad geológica imponente. Las zonas se producen en el proceso de apertura de las cuencas oceánicas. Posteriormente quedan implicadas en reajustes enérgicos.

Tal vez la primera clave sobre la naturaleza de las zonas de fractura la aportara el descubrimiento de que la cresta de la dorsal mesoceánica era un lugar de intensa actividad volcánica y sísmica y de gran flujo de calor procedente del interior de la tierra. Esos y otros hallazgos recibieron explicación en el marco de la teoría de la tectónica de placas, uno de cuyos principios establece el crecimiento de nueva corteza oceánica a lo largo de la cresta de las dorsales mesoceánicas, corteza que se agrega a las grandes placas que constituyen el fondo de las cuencas oceánicas. Las placas se llevan corteza de las crestas a velocidades de uno a veinte centímetros por año. Por tanto, las crestas constituyen los ejes de expansión del fondo oceánico. En el caso del Atlántico y otros, el movimiento de las placas oceánicas provoca la separación de continentes, brindando un mecanismo para la deriva continental.

Sin embargo, el significado geológico de los desencajamientos laterales de las crestas no se esclareció hasta 1965, cuando J. Tuzo Wilson, de la Universi-





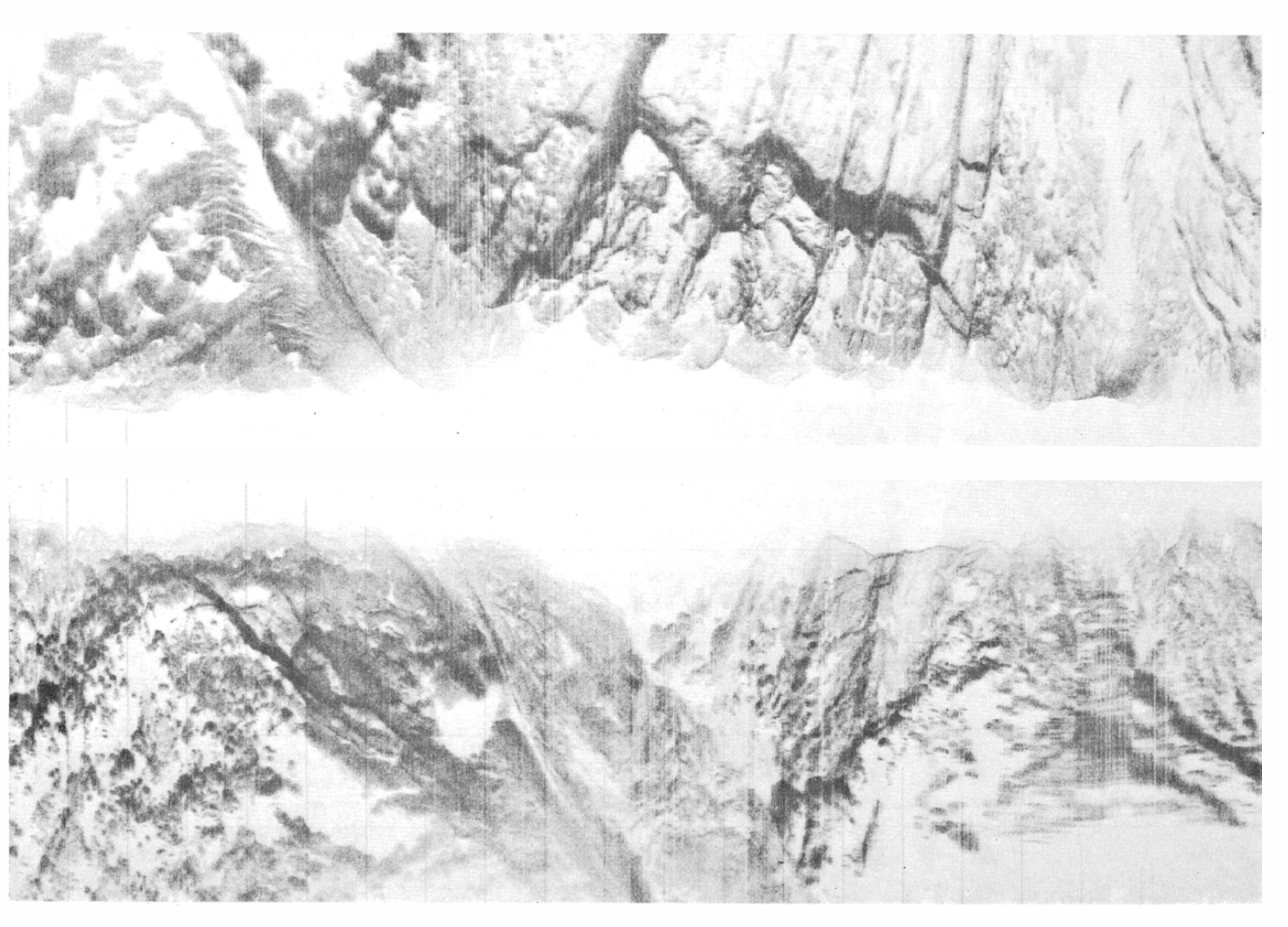
1. ZONA DE FRACTURA de Clipperton, que raja el fondo del Pacífico y produce un desencajamiento de 80 kilómetros en la prominencia submarina llamada dorsal del Pacífico oriental; la dislocación la determina la parte de la zona conocida

dad de Toronto, introdujo el concepto de falla transformante en la hipótesis de la expansión del fondo del mar. En ese marco, una placa oceánica se lleva corteza por un lado de la cresta de la dorsal y otra placa traslada corteza por el otro lado. Por consiguiente, a lo largo de un desencajamiento de dos segmentos de eje de dorsal, los bloques de corteza resbalarían uno frente a otro como si se desplazaran en sentidos opuestos. El desencajamiento es una falla transformante: una zona de lo que los geólogos llaman movimiento de salto según la dirección, o desgarre.

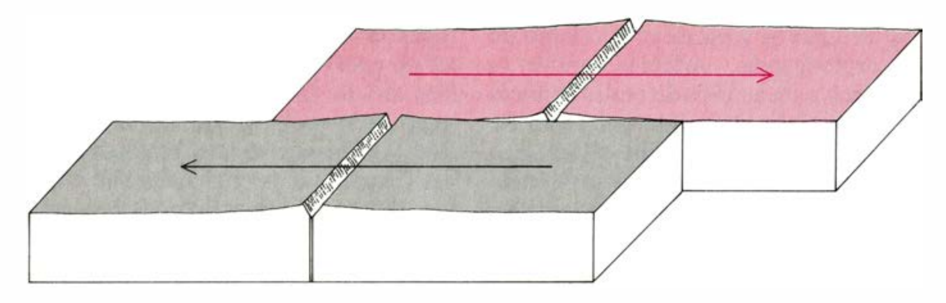
En el modelo de Wilson, el movimiento de desgarre se produce sólo a lo largo de la longitud de la transformante, y no más allá de ella, a lo largo de la línea que representa la prolongación de la falla transformante por la cuenca oceánica. Esa predicción la corrobora la observación de que en la transformante se concentran los epicentros de los terremotos y escasean a lo largo de su prolongación. Además, apoya la predicción un análisis de ondas sísmicas de terremotos efectuado por Lynn R. Sykes, del Observatorio Geológico Lamont-Doherty, de la Universidad de Columbia. Las ondas revelan la dirección en que resbalan súbitamente bloques corticales, generando terremotos en la zona de transformante. Probablemente la dirección de deslizamiento sea también la del movimiento tectónico más lento. El análisis de Sykes indica que, efectivamente, el movimiento es de salto según la dirección y que sigue las direcciones que Wilson predijo.

Sin duda, el modelo de Wilson de fallas transformantes es correcto en sus grandes líneas. Con todo, investigaciones recientes inducen a pensar que se trata de una simplificación excesiva de la realidad. Por una parte, el surco y las dorsales transversales que constituyen la rúbrica de fallas transformantes de gran dislocación pueden seguirse más allá de la zona desencajada, a lo largo de la prolongación de la dislocación. Ciertamente, la rúbrica de algunas transformantes del Atlántico ecuatorial puede seguirse de un lado a otro del Atlántico, es decir, desde la costa africana hasta la sudamericana. Junto al margen de cada continente está enterrada bajo sedimento, pero las prospecciones sísmicas revelan su presencia. La idea de que las fallas transformantes son lugares donde los bloques corticales se deslizan uno frente a otro más o menos pasivamente pierde terreno ante la que sostiene que son lugares de compleja actividad geológica de importancia fundamental para determinar la estructura y evolución de las cuencas oceánicas.

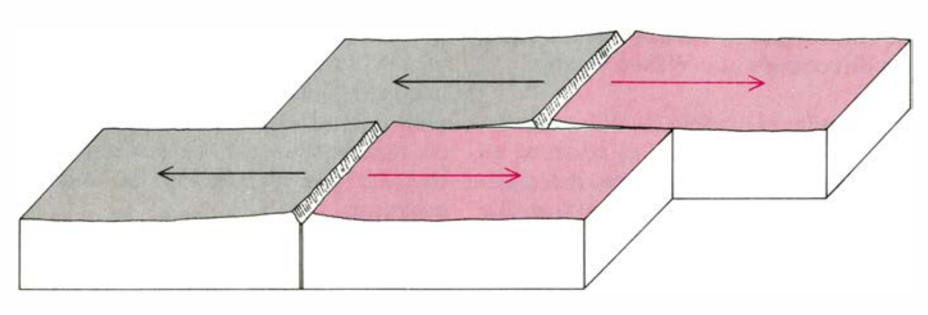
¿Qué determina la presencia de un surco profundo a lo largo de una falla transformante? Quizás entren en juego



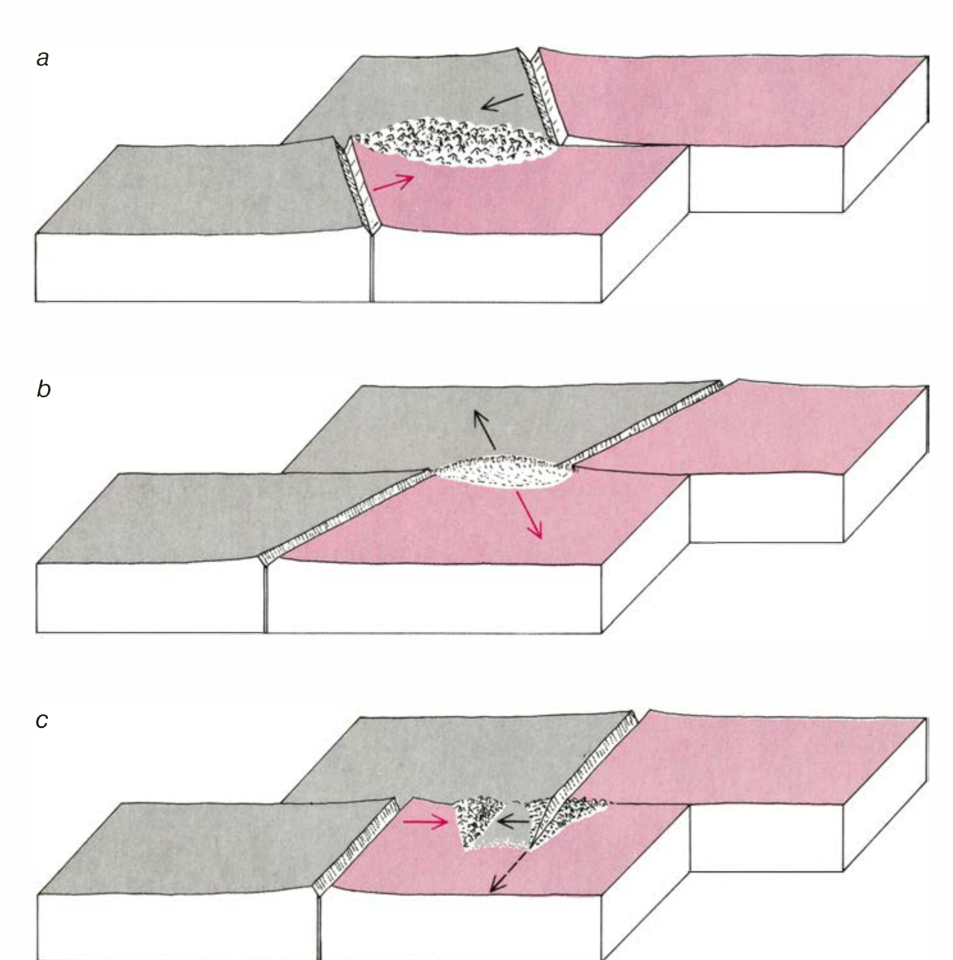
por falla transformante de Clipperton, que atraviesa la imagen diagonalmente, de arriba abajo. (Se distingue justo arriba como una banda prominente.) El Sea MARC, un aparato de sonar puesto a punto por uno de los autores (Crane) en el Observatorio Geológico Lamont-Doherty, de la Universidad de Columbia, obtuvo la imagen. Arrastrado de 200 a 500 metros sobre el fondo del mar, el Sea MARC dirige una banda de ondas acústicas a izquierda y derecha de su trayecto. En una imagen de Sea MARC, los puntos oscuros representan reflexiones fuertes de la banda, producidas, por ejemplo, por vertientes de riscos submarinos. La imagen reproducida aquí muestra un tramo de fondo marino de unos cinco kilómetros de anchura por 10 de longitud.



2. SEGUN LA PRIMERA HIPOTESIS sobre las zonas oceánicas de fractura éstas se asimilarían a fallas continentales de desgarre: líneas a lo largo de las cuales dos bloques de corteza resbalan frente a sí.



3. SEGUN LA TEORIA DE LA TECTONICA DE PLACAS, a lo largo de cada segmento de una dorsal se forma corteza oceánica. Así, la dorsal es estacionaria y a cada lado de ella diverge placa. Dos placas resbalan una frente a otra sólo a lo largo del desencajamiento entre los dos segmentos de eje de dorsal.



4. APARECEN TENSIONES EN UNA ZONA DE FRACTURA cuando las placas cambian su dirección de expansión. Si la dirección cambia de manera que las placas convergen en la zona de fractura (*a*), la corteza se comprime y provoca el ascenso del fondo del mar. Si la dirección cambia de modo que las placas divergen de la zona (*b*), la corteza se estira y debilita, formándose una "cuenca de estiramiento". La propagación de uno de los segmentos del eje de la dorsal (*c*) también puede generar compresión.

varios factores. Fundamentalmente, la corteza oceánica nueva se forma a lo largo del eje de una dorsal mesoceánica porque la emanación de magma basáltico caliente, producido en el manto terrestre a unos 30 o 60 kilómetros por debajo del fondo oceánico, se enfría y solidifica en el eje de la dorsal al llegar al fondo oceánico. El nivel alcanzado por el magma corresponde a la profundidad del eje (es decir, a la distancia al nivel del mar) y lo determinan la cantidad de magma, su temperatura y viscosidad y el arrastre debido a la interacción del magma con las paredes relativamente frías del conducto por donde se produce la emanación. La corteza se aleja del eje y, en su desplazamiento, sigue enfriándose y contrayéndose lentamente. En consecuencia, la profundidad del fondo oceánico aumenta con la edad del fondo y con su distancia al eje donde se formó. La relación la expresa una ley gratificadoramente sencilla, verificada en los tres océanos principales, Atlántico, Pacífico e Indico: la profundidad del océano aumenta con la raíz cuadrada de la edad de la corteza.

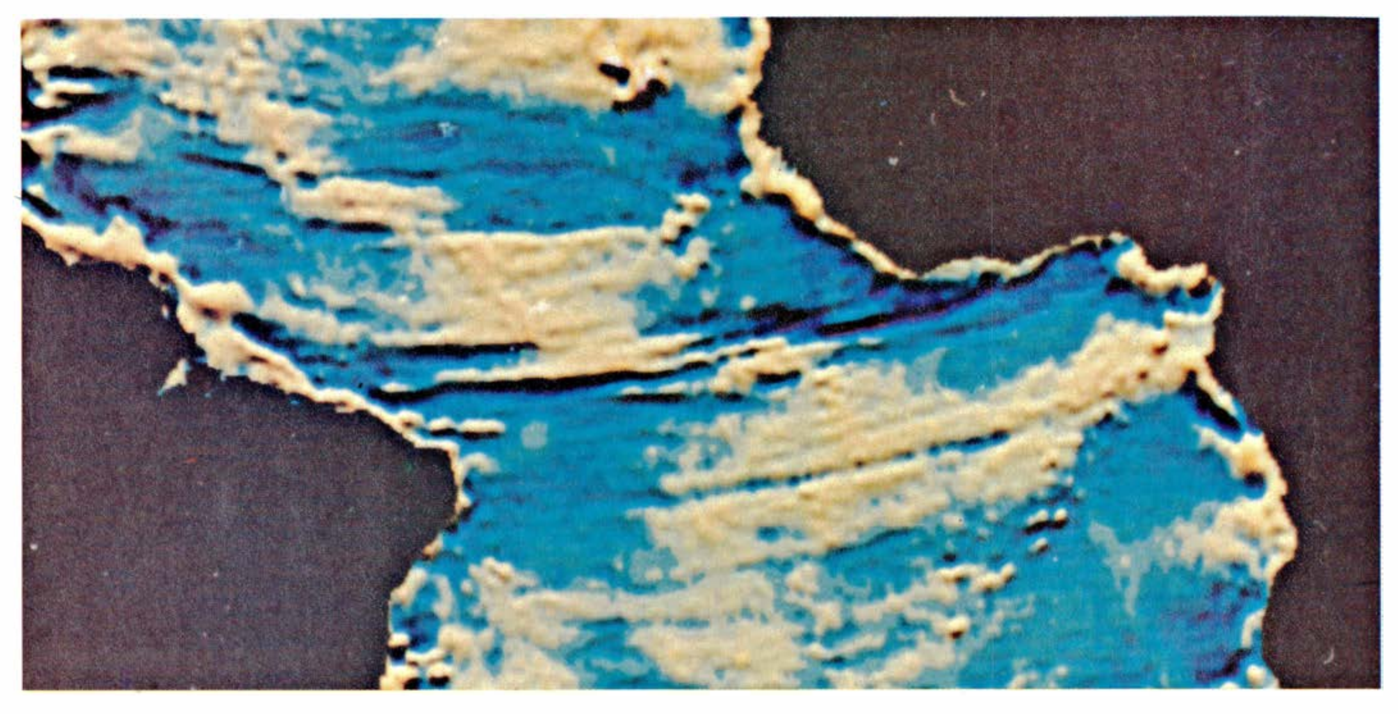
En una falla transformante esta ley simple se rompe: el surco es, en esencia, una profundidad oceánica anómala. Pero en una falla transformante un segmento de eje de dorsal caliente y joven se junta con litosfera más fría y antigua. (La litosfera incluye la corteza y la parte superior, relativamente rígida, del manto. Juntas, determinan el grosor de una placa tectónica.) Así pues, la emanación de material caliente a lo largo del eje cerca de una falla transformante debe interactuar con una superficie fría que no encuentra en otras partes. Debe entonces "congelarse" en un nivel más profundo, produciendo un bajo topográfico en la intersección del eje con la transformante. La expansión de la placa extiende el bajo topográfico hacia un surco que puede atravesar toda la anchura de una cuenca oceánica. El surco puede incluso ensancharse durante la fase inicial de expansión del fondo del mar, porque el enfriamiento de una placa presumiblemente provoca su contracción, no sólo vertical, sino también horizontalmente.

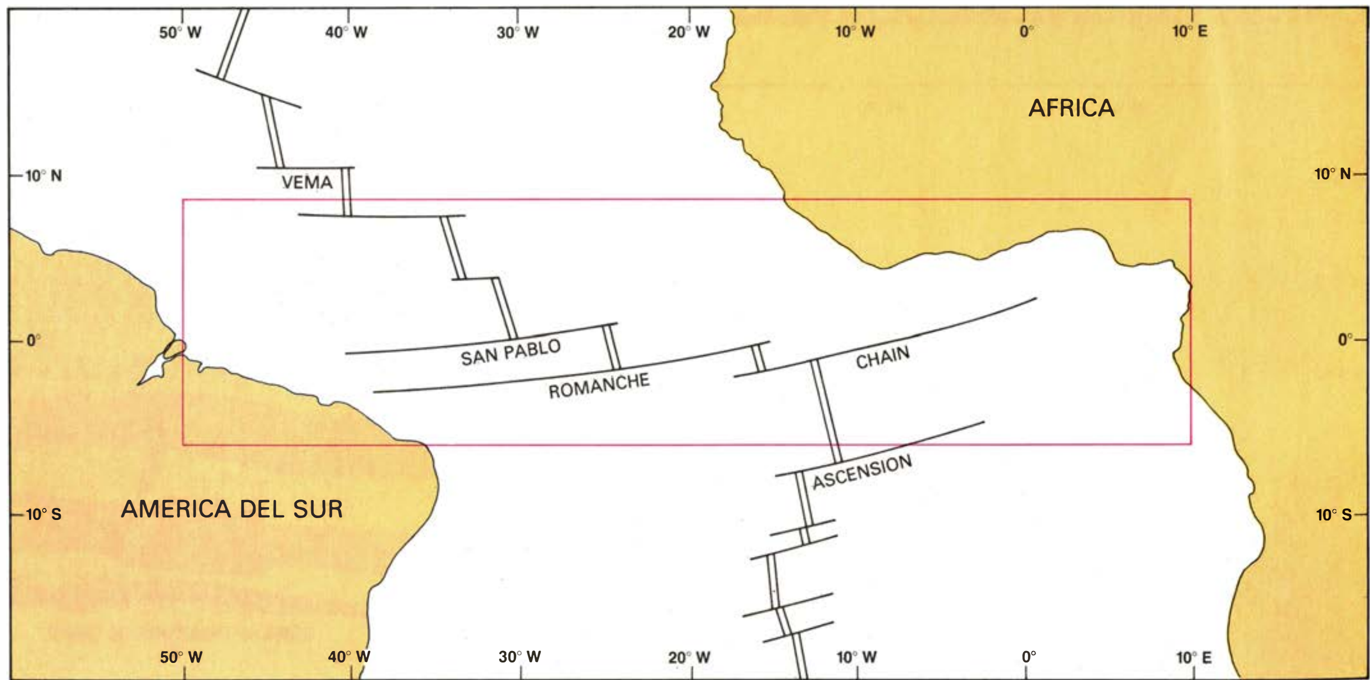
Las dorsales transversales que flanquean el surco constituyen otra violación de la ley que vincula la profundidad del océano a la edad de la corteza oceánica. Son cumbres anómalas. ¿Cómo se originan?

Durante la última década se han tomado muestras de las rocas que forman las paredes de surcos transformantes y las laderas de dorsales transversales. La mayoría de las muestras se obtuvieron

arañando el fondo del mar con una draga maniobrada desde un barco; algunas las extrajo el sumergible de investigación *Alvin* en el transcurso de un estudio de la falla transformante Oceanographer (en el que intervinimos los autores, junto con Paul J. Fox, de la Universidad de Rhode Island, e investigadores de la Universidad estatal de Nueva York en Albany y de la Institución Oceanográfica Woods Hole). La transformante Oceanographer se halla a 35 grados de latitud norte, cerca de las Azores, en el Atlántico norte.

Entre las rocas extraídas abundaban de dos tipos. El primero, peridotitas: rocas constituidas principalmente por olivino, piroxenos y espinelas. A su vez, olivino y piroxenos son silicatos de elementos tales como magnesio y hierro; las espinelas son óxidos de magnesio, aluminio y cromo. Por tanto, el contenido en magnesio de una peridotita (expresado como contenido neto en óxido de magnesio) es alto: puede llegar al 40 por ciento. El contenido en magnesio de un basalto (la roca característica de la corteza oceánica) es de sólo una cuarta parte. Físicamente, la peridotita es una roca verdosa oscura cuya densidad, grande, se cifra en 3,2 gramos por centímetro cúbico. Tal densidad concuerda con la hipótesis de que las peridotitas son los constituyentes principales del manto superior. La den-

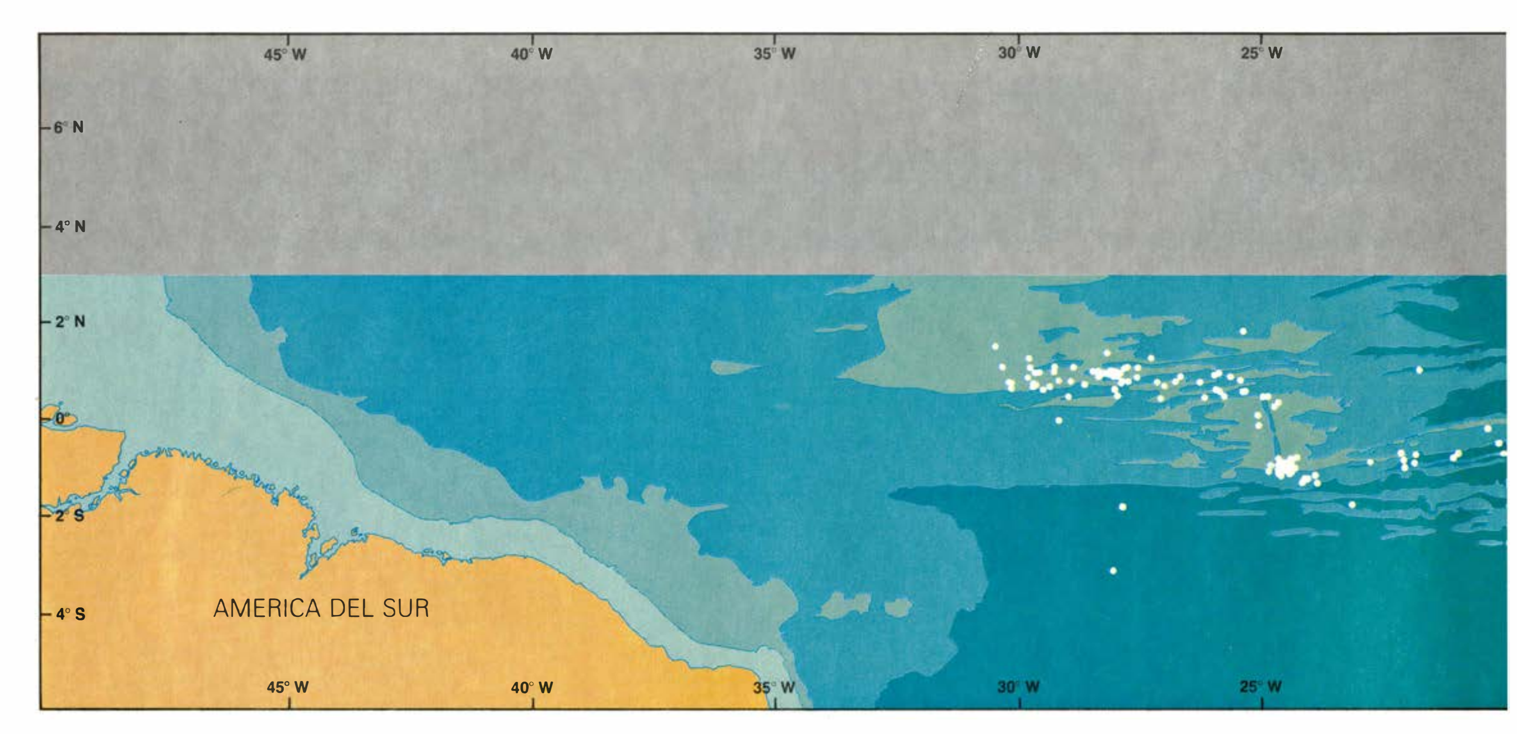




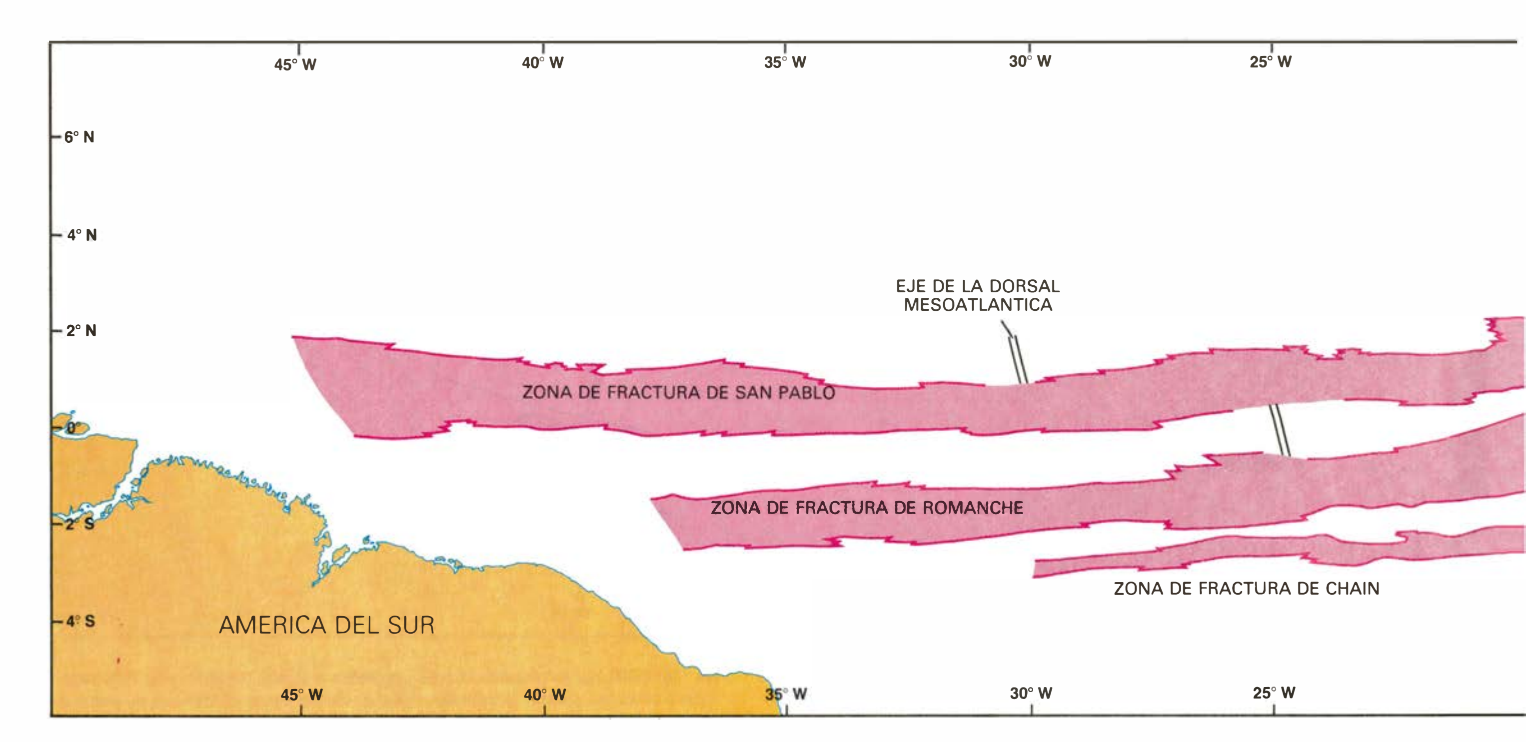
5. MEDICIONES DEL NIVEL DEL MAR, efectuadas por radar desde un satélite, proporcionaron esta imagen de una serie de zonas de fractura que cortan la dorsal mesoatlántica. Las manchas blancas señalan regiones, como la propia dorsal, donde el suelo oceánico no es muy profundo. Allí el nivel del océano sube. Las manchas oscuras muestran profundidades crecientes, entre las que destacan profundos surcos asociados a zonas oceánicas de fractura. Allí el nivel del océano es inferior. Muchas de las zonas atraviesan de parte a parte el Atlántico ecuatorial. Las zonas se identifican en el mapa anejo. Un rectángulo señala la extensión que cubren las figuras 6 y 7. La imagen la proporcionó William F. Haxby, del Observatorio Geológico Lamont-Doherty.

sidad del manto superior se ha deducido, de hecho, de la velocidad a la que se propagan a su través las ondas sísmicas. El resultado sugiere que la densidad del manto superior bajo los océanos varía entre 3,1 y 3,3 gramos por centímetro cúbico.

El segundo tipo de rocas abundantes eran los gabros. Mineralógica y químicamente se parecen mucho al basalto: constan principalmente de silicatos de calcio y aluminio, que forman minerales como el feldespato. También se parecen físicamente al basalto: son grisáceos y tienen una densidad que ronda los 2,8 gramos por centímetro cúbico. La diferencia principal entre un gabro y un basalto radica en que los cristales del gabro son mayores. Por ello se deduce que tanto gabros como basaltos se originan por la fusión parcial de peridotitas en el manto. Las partes de la roca fundida que llegan al fondo del mar se enfrían rápidamente al hacer erupción, rindiendo basalto. Las otras partes se enfrían lentamente a profundidades mayores; en este caso se produce gabro. El gabro solidifica bajo el eje de las dorsales, en cámaras magmáticas de la parte inferior de la corteza.

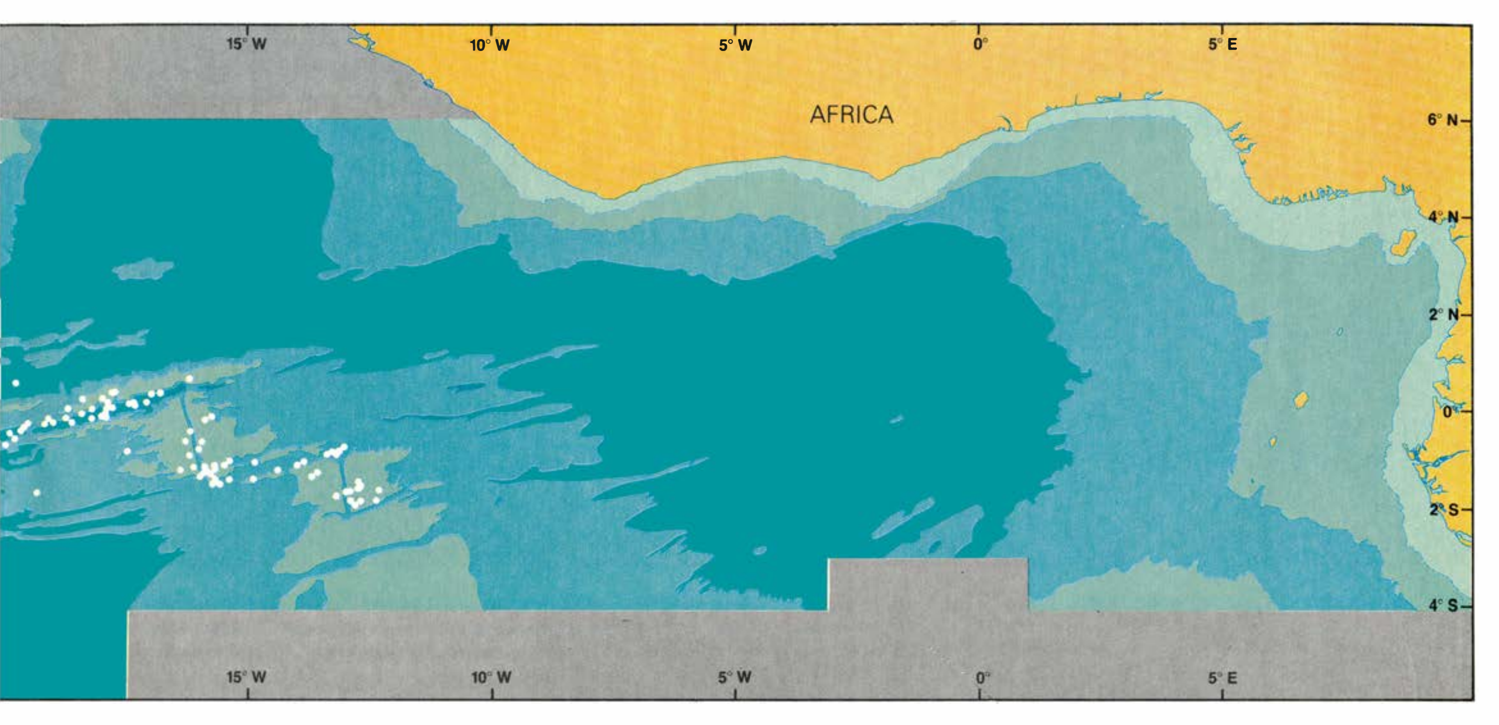


6. UN COMPLEJO BORDE DE PLACA, compuesto por segmentos de dorsal mesoceánica desencajados por una serie de fallas transformantes, se deduce de dos series de pruebas: la topografía del fondo oceánico y la distribución de epicentros de terremotos. Los segmentos de la dorsal son lugares donde el fondo marino se acaba de formar, razón por la cual está caliente y se halla expandido verticalmente; los segmentos de dorsal son, por ello, partes relati-

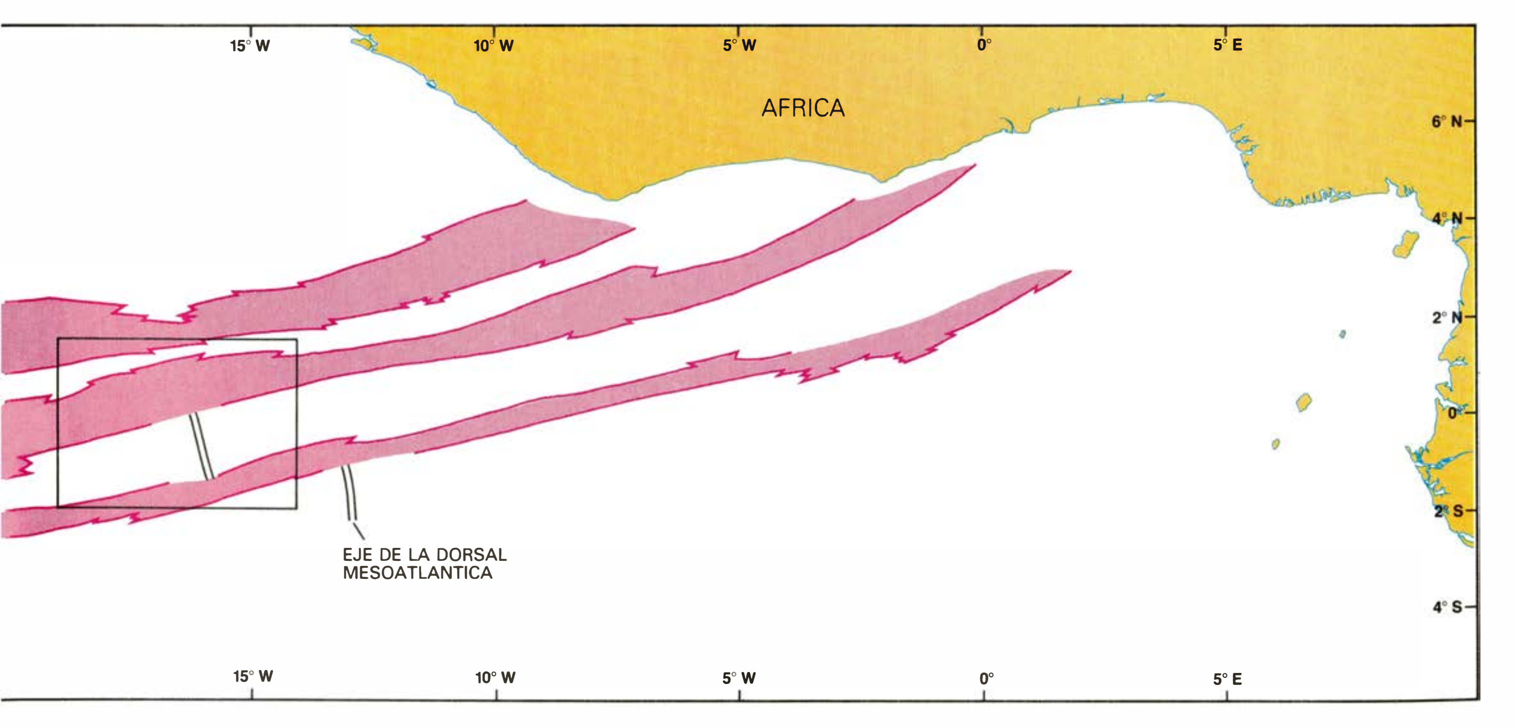


7. TRES ZONAS PRINCIPALES DE FRACTURA, representadas en rojo, explican los largos desencajamientos de la dorsal mesoatlántica que se distinguen en la figura 6. Cada zona comprende una falla transformante que discurre a lo largo de las dislocaciones. Más allá de los límites del desencajamiento, la zona se extiende a lo ancho del Atlántico, cual surco flanqueado por dorsales de fondo marino anómalo. Junto a las costas de Africa y Sudamérica, las

El punto decisivo estriba en que ambos tipos de roca abundantes en las fallas transformantes –peridotita y gabro– suelen ser de asentamiento profundo. La peridotita se halla en el manto superior y, el gabro, en la corteza inferior. Por consiguiente, la presencia de estas rocas sugiere que las dorsales transversales son astillas de corteza profunda y manto superior que, de alguna manera, han resultado levantadas. A este respecto, el estudio de secciones de dorsales transversales constituidas integramente por peridotita derivada del manto se ha ofrecido particularmente revelador. Los minerales de la peridotita sólo son estables a las altas temperaturas y presiones del manto. En cualquier otro contexto tienden a transformarse en otros minerales. Además, algunos de los minerales de la peridotita, en particular piroxenos y espinelas, varían levemente de composición química en función de la temperatura y la presión (es decir de la profundidad) a la que se forman. Las asociaciones minerales de las peridotitas muestreadas en fallas transformantes inducen a pensar que proceden de profundidades que superan los 30 kilóme-

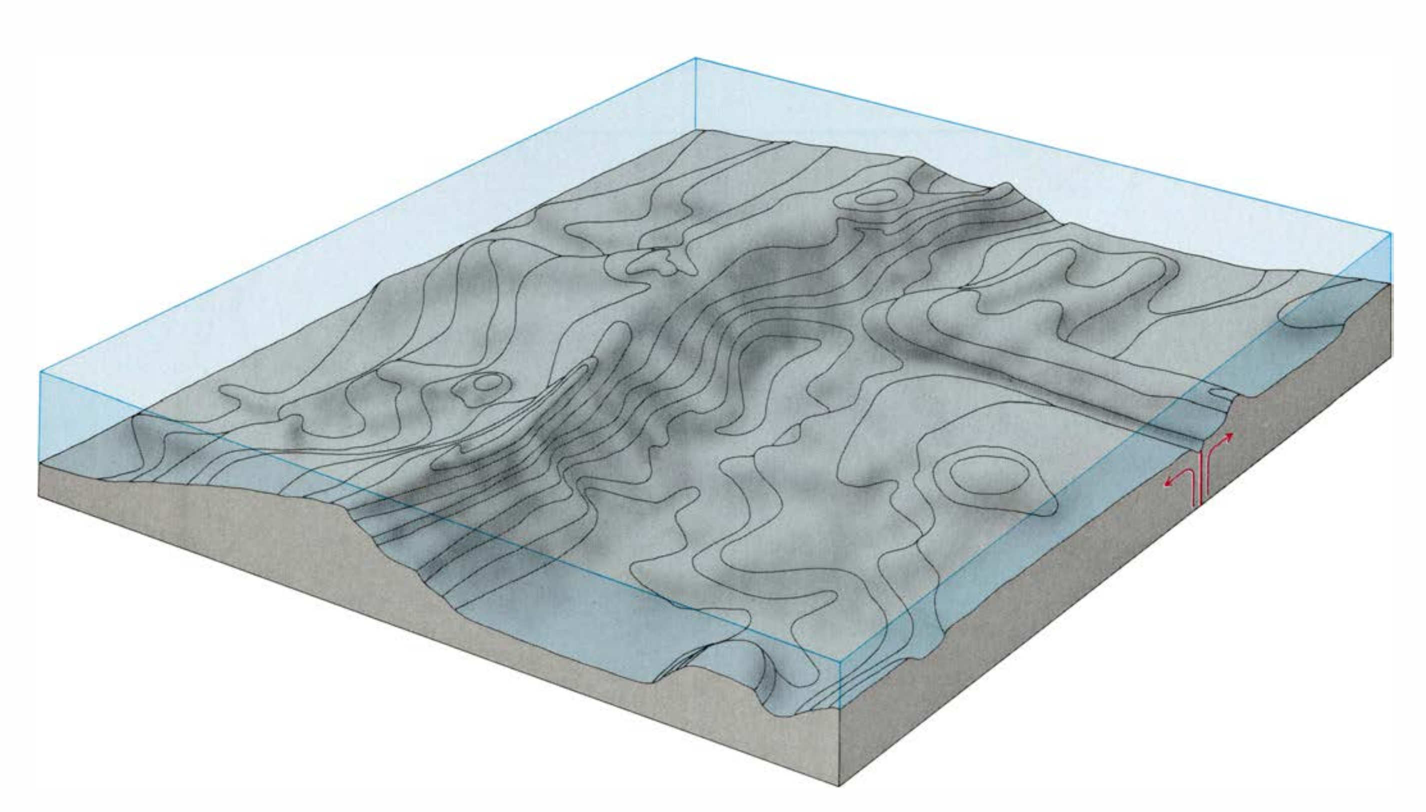


vamente someras de la cuenca del océano Atlántico. Las fallas transformantes son lugares de salto cortical súbito en el sentido de la expansión del fondo del mar; de ahí que, junto con los segmentos de eje de dorsal, constituyan sitios donde tienden a producirse seísmos. Las isóbatas del mapa indican profundidades de 1000, 3500 y 4500 metros. Los puntos blancos designan posiciones de terremotos registrados entre 1960 y 1970. Su distribución resigue las fallas.



zonas han quedado cubiertas por sedimento continental; aún así, los datos sísmicos confirman su presencia. Las tres zonas se hallan en el corazón de lo que se ha denominado zona de megacizalla ecuatorial. Sus posiciones las identificó Marcus Gorini, del Observatorio Geológico Lamont-Doherty. Las porciones de la zona de fractura de Romanche y de la dorsal mesoatlántica que se recuadran aquí se representan con mayor detalle en la figura 8.

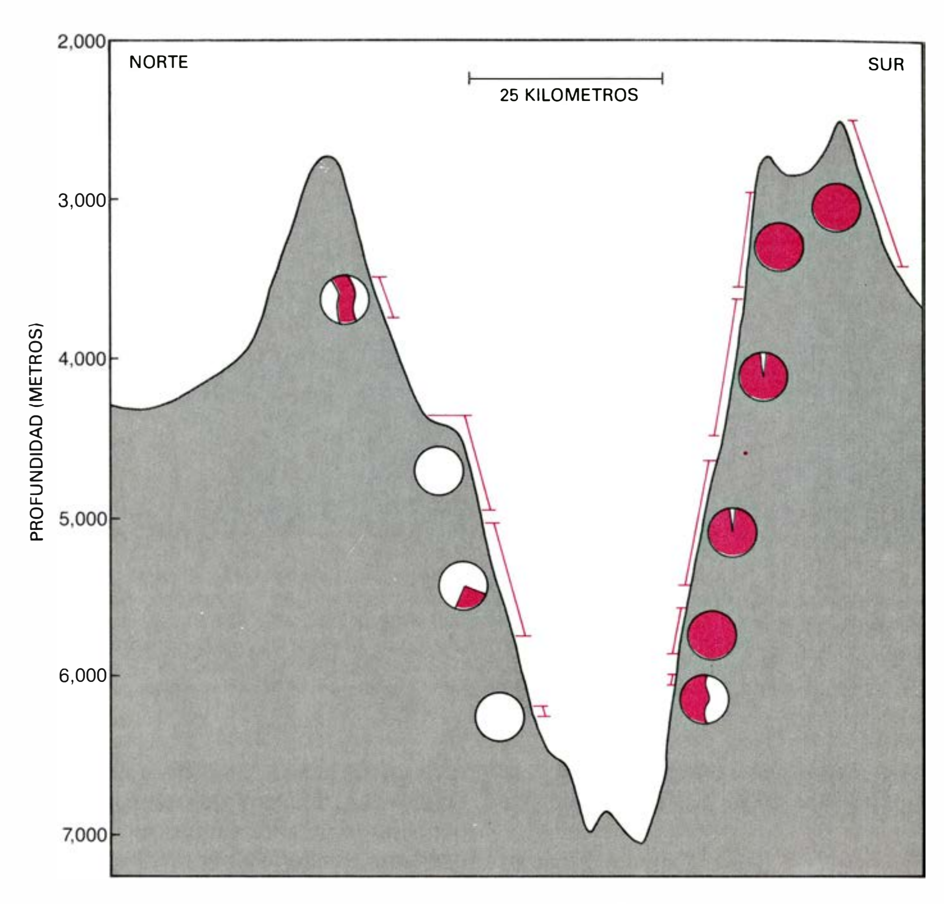
8. ZONA DE FRACTURA DE ROMANCHE, la más pronunciada de cuantas componen la megacizalla ecuatorial; su profundidad va de menos de un kilómetro, en la cima de sus prominencias, a más de 7,5 kilómetros, en las simas más profundas de sus surcos. He aquí una cartografía detallada del relieve topográfico de parte de la zona. Las isóbatas señalan intervalos de 500 metros, entre 1000 y 5000 metros, con curvas adicionales a 6000 y 7000 metros. El modelo es claro: un surco patente flanqueado por prominencias visibles. Un segmento de la dorsal mesoatlántica se une a la zona desde el sur.



9. PERSPECTIVA OBLICUA de parte de la zona de fractura de Romanche, que destaca su accidentada topografía, exagerando la escala vertical. Isóbatas a intervalos de 500 metros, desde 1000 a 5000 metros. Según la teoría de la tectónica de placas, la elevación del fondo marino disminuye con la edad de la corteza. Los surcos y las prominencias que caracterizan una zona de fractura oceánica son, pues, anomalías de la habitual relación edad-profundidad.

tros. (El límite entre la corteza y el manto generalmente queda a una profundidad de cuatro a cinco kilómetros.) De alguna manera, tales peridotitas han remontado esa distancia en forma de cuerpos esencialmente sólidos.

En el curso de su ascenso, la peridotita habrá llegado a niveles de la corteza, a pocos kilómetros bajo el fondo del mar, donde se encuentra agua marina. Resulta ciertamente verosímil que penetre agua marina en la corteza en dorsales mesoceánicas y fallas transformantes, porque en esos lugares la corteza está sumamente fracturada y, por ello, es permeable. Algunos minerales de la peridotita (en especial olivino y piroxenos) pueden reaccionar con el agua marina, y así se transforman gradualmente en otros minerales, en particular en la serie de silicatos de magnesio hidratados llamados serpentinas. Consecuentemente, la densidad de la peridotita disminuye de manera espectacular, de 3,2 a 2,6 o incluso menos. La reducción debe tender a facilitar el ascenso de la roca a través de los materiales más densos de la corteza, incluidos gabro y basalto.



10. PERFIL TRANSVERSAL de Romanche. Una composición circular en cada lugar indica la identidad de las rocas. Algunas (*en blanco*) eran normales, constituidas por basalto, la roca de la que está hecha la corteza oceánica. En cambio, se reveló que la peridotita (*en color*) formaba 4,5 kilómetros de la altura de la pared sur del surco de Romanche. La peridotita no suele formar parte de la corteza oceánica.

No se afirma con ello que el levantamiento de roca de emplazamiento profundo en zonas oceánicas de fractura se haya deducido sólo mediante análisis sutiles de rocas extraídas de grandes profundidades del océano con dragas o sumergibles de investigación. Por lo menos en un caso, las cimas de un cuerpo procedente del manto sobresalen del nivel del mar: las islitas Santos Pedro y Pablo ascienden desde el profundo fondo oceánico en medio del Atlántico, inmediatamente al norte del ecuador. En 1831, Charles Darwin, a bordo del *Beagle*, recaló brevemente en su terreno desolado y azotado por el oleaje. Apreció algo geológicamente inusual: no parecían volcánicas. Hoy se sabe que están hechas de peridotitas. Son un fragmento de manto superior levantado cerca de la intersección de la transformante de San Pablo con el eje de la dorsal mesoatlántica.

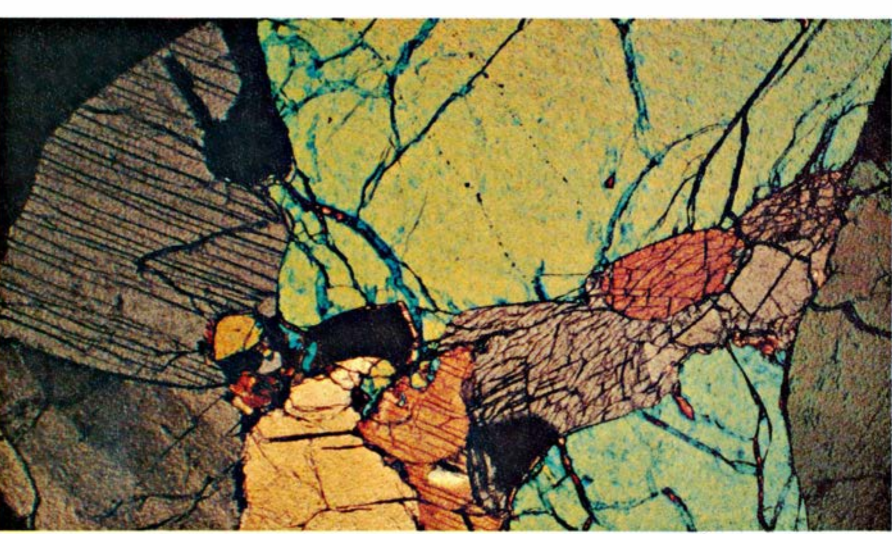
Además, se ha constatado que partes de otras dorsales transversales del Atlántico ecuatorial han emergido en forma de islas en épocas recientes, geológicamente hablando. En la transformante Vema, a 11 grados al norte del ecuador, el eje de la dorsal mesoatlántica está dislocado unos 300 kilómetros. A todo lo largo del desencajamiento, una astilla levantada de corteza oceánica queda en el lado sur del surco de transformante. La cima de la astilla levantada queda a sólo 600 metros del nivel del mar. Un muestreo detallado de la astilla revela que está coronada por calizas arrecifales parecidas a las que rodean islas como las Bahamas.

Varias series de pruebas han ayudado a deducir el ambiente en que se formaron las calizas. Por una parte, las calizas comprenden estructuras de agua somera, por ejemplo, oolitos: depósitos calcáreos formados cuando agua fría fluye hacia bajos cálidos y someros. Contienen restos fósiles de corales, algas calcáreas y moluscos, que viven en aguas someras o en la zona comprendida entre pleamar y bajamar. Además, las calizas muestran alteraciones típicas de erosión en tierra. Finalmente, las proporciones de los isótopos de oxígeno son reveladoras. Si un organismo marino forma una concha de carbonato cálcico, los átomos de oxígeno del CO_3Ca tendrán un cociente oxígeno 16/oxígeno 18 relacionado con el del oxígeno disuelto en el agua marina. Si, posteriormente, la concha se expone al agua de lluvia, en la que la razón entre isótopos es levemente diferente, ésta cambiará.

Las observaciones confirman que las calizas se formaron cerca del nivel del mar como parte de arrecifes y lagunas coralinos. Se ha comprobado también que las calizas se meteorizaron y alteraron su composición química subaéreamente. Ello significa que las partes más someras de la dorsal transversal de Vema emergieron en una o varias islas en algún momento del pasado. ¿Cuándo? La identificación de los caparazones fósiles de foraminíferos, organismos marinos microscópicos, sugiere que la dorsal transversal de Vema estuvo sobre el nivel del mar, continua o intermitentemente, desde mediados del Mioceno hasta mediados del Plioceno, es decir, desde hace unos 10 hasta hace unos tres millones de años. Desde entonces se ha ido hundiendo en el mar a un ritmo medio de 0,3 milímetros por año.

La situación es similar en la intersección oriental de la transformante de Romanche con la dorsal mesoatlántica. Aquí, la dorsal transversal de Romanche se acerca a la superficie hasta llegar a poco más de 1000 metros del nivel del mar a lo largo de un tramo de más de 100 kilómetros. De nuevo pudimos demostrar que la cima de la dorsal estaba coronada por un arrecife fósil y que estuvo emergida hace unos cinco millones de años como isla o archipiélago. Ha



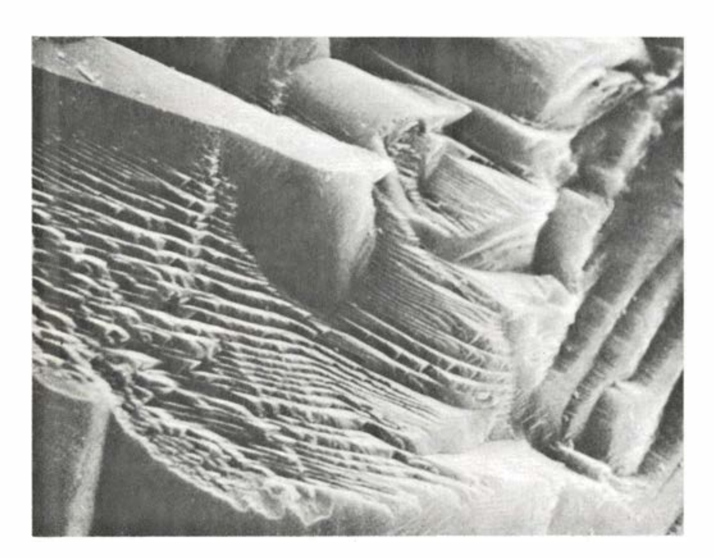


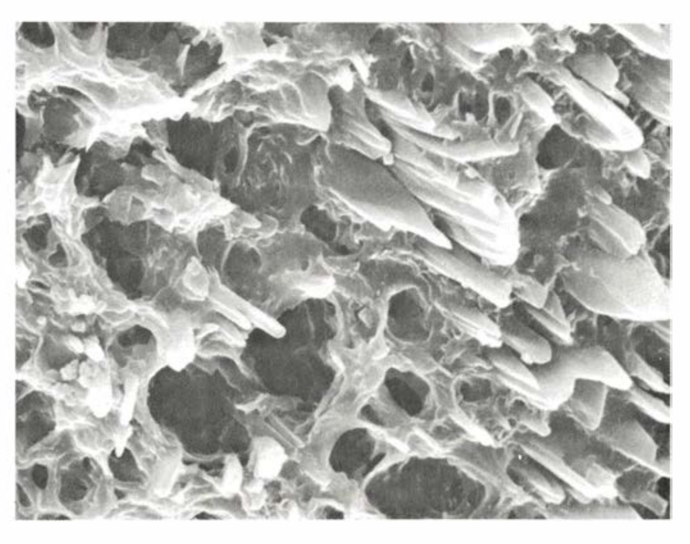
11. PERIDOTITA de la zona de fractura Oceanographer, en el Atlántico norte, fotografiada desde el sumergible de investigación *Alvin* (*izquierda*); su estructura interna puede deducirse de una peridotita de la zona de fractura del mar Rojo, fotografiada en lámina delgada iluminada con luz polarizada (*derecha*). La peridotita consta de tres minerales: piroxeno, olivino y espinela. Arriba, a la izquierda, aparece, en micrografía de Enrico Bonatti, un cristal de piroxeno (*pardo en esta iluminación*); cristales de olivino (*verdes y anaranjados*) y de espinela (*negros*) llenan el resto del campo. La porción ampliada abarca unos 0,5 centímetros de longitud. Se cree que la peridotita ocupa la parte superior del manto, la capa subyacente a la corteza. Por tanto, su presencia en las partes elevadas de zonas de fractura sugiere que las elevaciones son astillas de corteza y manto superior levantadas en vertical hasta 30 km.

venido subsidiendo desde entonces a un ritmo medio de 0,2 milímetros por año.

Aun a riesgo de simplificar en demasía, consideremos muy esquemáticamente qué ocurriría en una falla transformante oceánica si la dirección de expansión de las placas oceánicas que resbalan una frente a otra a todo lo largo de una transformante cambiara, siquiera levemente, respecto de la orientación según la cual la transformante forma un ángulo recto con cada segmento del eje de la dorsal [*véanse las figuras 2, 3 y 4*]. Supongamos que los ángulos aumentan. Las placas ya no resbalarían una frente a otra, sino que divergerían. La zona de transformante quedaría sometida a extensión; la zona, por tanto, tendería a abrirse. Por ejemplo, la zona podría ensancharse, y por la parte extendida y debilitada de la corteza oceánica podría subir algo de magma basáltico. Tal transformante se ha denominado, en consecuencia, "rezumante". Por el contrario, los ángulos podrían disminuir. Entonces, las placas convergerían en la zona de transformante. La corteza de la zona se deformaría por compresión lateral, y cabe suponer que los movimientos verticales resultantes determinaran el levantamiento tectónico de astillas de corteza.

¿Existe alguna prueba de que la dirección de expansión de las placas oceánicas haya cambiado en el transcurso del tiempo? Hace más de una década Henry W. Menard, Jr., y Tanya M. Atwater, de la Institución Scripps de Oceanografía, señalaron que las anomalías magnéticas de varias partes de las cuencas oceánicas del globo mostraban cambios de orientación. Las anomalías magnéticas tal vez sean la prueba más importante en favor de la teoría de la tectónica de placas. Se originan al solidificarse corteza oceánica por enfriamiento de la lava; el campo magnético de la lava queda alineado según la orientación del campo terres-





12. MICROGRAFIAS DE BARRIDO de peridotitas procedentes de zonas de fractura oceánicas, que muestran su alteración en contacto con agua marina. La peridotita de la izquierda procede del mar Rojo y no está alterada: la superficie de un cristal de piroxeno está escalonada en una serie de exfoliaciones. La peridotita de la derecha procede de Romanche, y está muy alterada: el contacto con agua marina que se filtra por grietas hacia el interior de la corteza oceánica ha hidratado su piroxeno, formando una malla de cristales del mineral serpentina. La hidratación y la porosidad de la malla reducen notablemente la densidad de la peridotita, lo que tal vez haya facilitado su ascenso desde el manto superior. Microfotografías de Dee Breger, de Lamont.

tre en ese momento. El campo terrestre ha invertido su polaridad muchas veces en el transcurso de la historia geológica; por consiguiente, el campo de la corteza formada en un eje de dorsal mesoceánica, y separada de él por expansión del fondo marino, debe llevar impreso el registro de una serie de inversiones. Así es. Más aún, el patrón de inversiones a un lado de un eje de dorsal mesoceánica es la imagen especular del patrón del otro lado del eje. Menard y Atwater apreciaron que la orientación de las "bandas" magnéticas entre inversiones sucesivas no siempre corría paralela a las orientaciones de las dorsales mesoceánicas. De lo que se infiere que la dirección de expansión del fondo del mar puede cambiar.

Las propias zonas oceánicas de fractura aportan pruebas de que la dirección de expansión del fondo del mar puede cambiar. Los surcos y las dorsales transversales producidas en fallas transformantes y transportadas a través de las cuencas oceánicas por expansión del fondo marino revelan abruptos cambios de orientación. Esos cortes súbitos se distinguen con claridad en los mapas del fondo del Atlántico cartografiados mediante altimetría de radar por William F. Haxby, colega nuestro en Lamont-Doherty. Presumiblemente, cada interrupción la provocaron las tensiones originadas cuando las placas cambiaron su dirección de movimiento.

Así pues, cualitativamente, podríamos afirmar que los movimientos tectónicos verticales que generan los bajos anómalamente someros en zonas de transformantes los producen, sobre todo, esfuerzos compresivos resultantes de la reorientación de la falla transformante en respuesta a cambios de dirección del movimiento de una placa. No se cuenta todavía con datos suficientes para cuantificar tal aseveración. Para ello habrá que efectuar mediciones de la deformación de la corteza y emprender una extracción generalizada de muestras de rocas.

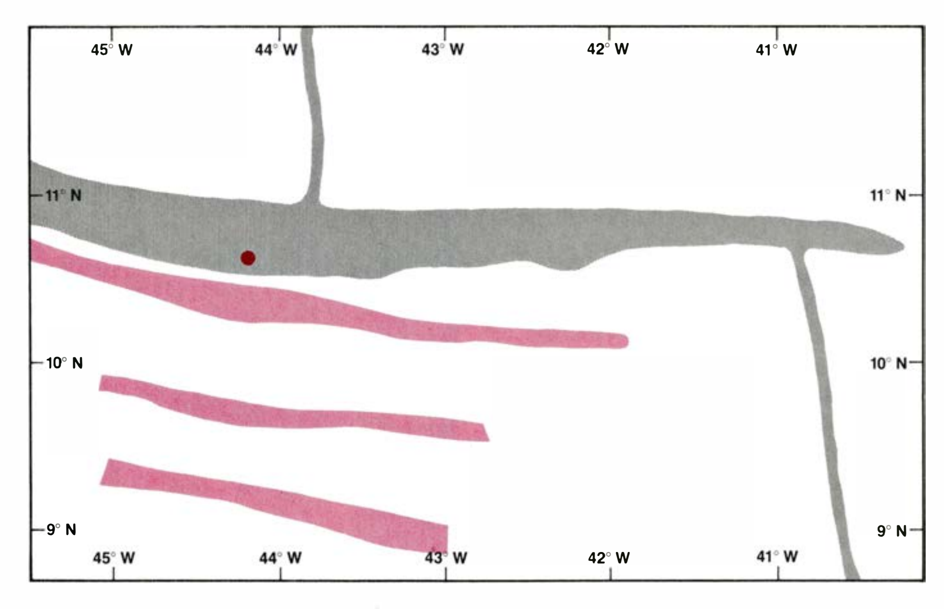
Donde mejor podemos hoy aproximarnos a un análisis cuantitativo es en la zona de fractura de Vema. A principios de la década de 1970, T. H. van Andel, de la Institución Scripps, se percató de la presencia de un juego de valles inactivos de transformante al sur de la transformante que hoy se muestra activa. Los valles inactivos forman un ángulo de entre 10 y 15 grados con respecto a la transformante activa. Además, esos valles inactivos sólo se distinguen en corteza oceánica de más de 10 millones de años de antigüedad.

Estas observaciones encuentran su explicación en el marco del siguiente modelo. Hace más de 10 millones de años, una falla transformante se mostraba activa al sur de la actual transformante de Vema. El extremo de uno de los segmentos del eje de dorsal mesoatlántica se unió con la falla y luego se propagó hacia el norte. Una propagación longitudinal de ese tipo la han documentado en el Pacífico oriental, el golfo de Adén, el mar Rojo y otras zonas, investigadores diversos, entre los que se cuentan R. H. Hey, de la Institución Scripps, y Vincent Courtillot, de la Universidad de París. Después, la transformante Vema se reubicó, de modo que la falla y los segmentos de eje que ella une adquirieron la geometría que hoy se observa: la falla se junta con los segmentos de eje en ángulos rectos y las tensiones se relajan.

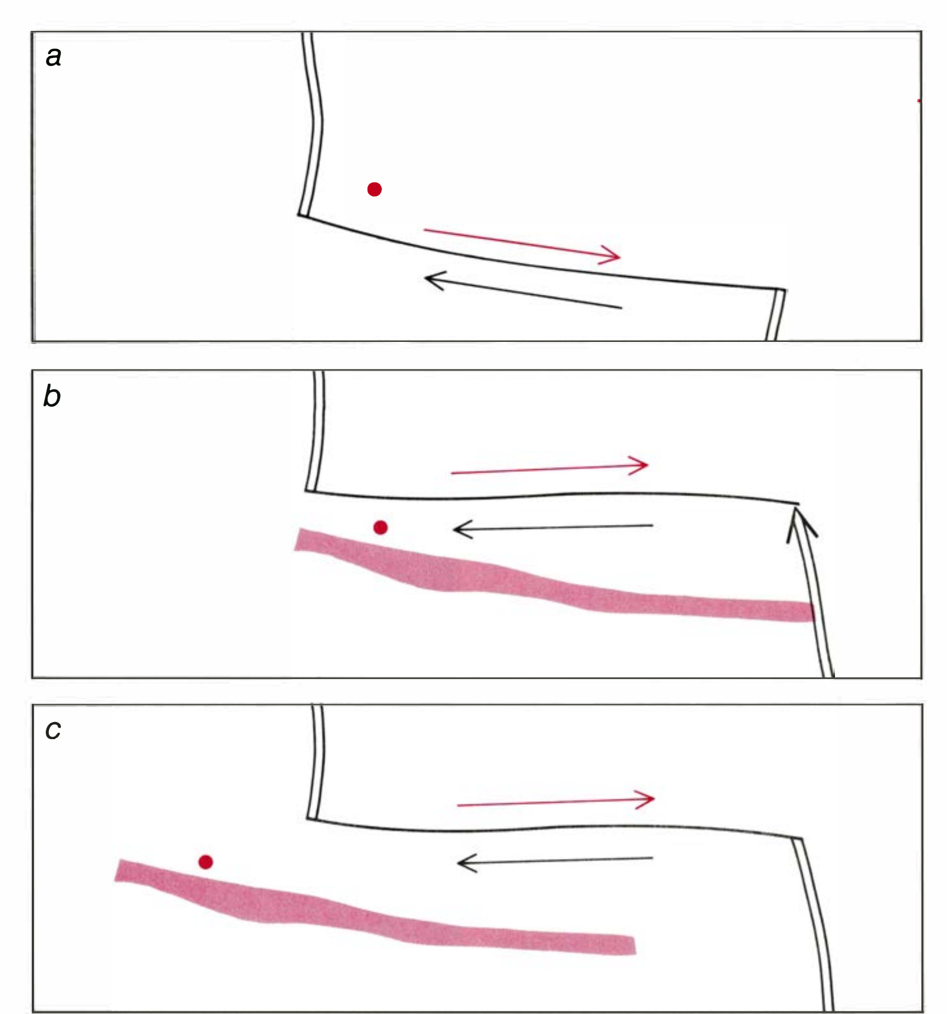
El modelo explica un curioso descubrimiento que hicimos en Vema: el bloque cortical levantado del lado sur de la actual falla transformante es anómalamente antiguo. En particular, parte de la caliza extraída del bloque tiene más de 55 millones de años. (Es caliza que no corresponde a un arrecife fósil, sino a la solidificación de sedimento calcáreo en el fondo oceánico.) Sin embargo, la tasa de expansión del fondo marino en Vema apenas rebasa el centímetro por año. Si la corteza basáltica subyacente a la caliza se había formado en el segmento de eje de dorsal situado al sur de la actual transformante de Vema, desplazándose, cubierta de caliza, hasta su posición actual, la caliza no tendría más de 30 millones de años.

Sugerimos que el bloque cortical coronado de caliza ha experimentado un proceso de expansión oscilatoria. Antes de la reubicación de la transformante de Vema, la caliza se hallaba al norte de la falla, y se movía hacia el este formando parte de la placa africana. Después de reubicarse (en la disposición actual) queda al sur de la falla y se desplaza al oeste como parte de la placa americana. En virtud de la inversión de su dirección de desplazamiento, ha estado moviéndose durante más de 55 millones de años, sin abandonar la región. Es probable que la transición de la falla de una configuración a la otra se acompañara de las fuertes tensiones compresivas que levantaron el bloque cortical en el que se encuentra la caliza.

Las observaciones y conclusiones extraídas de la zona de fractura de Vema demuestran que las largas fallas transformantes que segmentan una



13. MAPA DE LA ZONA DE FRACTURA DE VEMA, en la megacizalla ecuatorial, que muestra la presencia de numerosos valles (presumiblemente fallas transformantes inactivas) situados al sur de la transformante que hoy es activa. El mapa resume el trabajo de T. H. van Andel, de la Institución Scripps de Oceanografía. En un sitio de la zona de Vema (*punto rojo*) la roca era caliza de 55 millones de años de antigüedad. Si la caliza hubiera estado siempre al sur de la transformante de Vema, y alcanzado su posición actual por expansión del fondo del mar, hacia el oeste, sólo tendría la mitad de dicha edad.



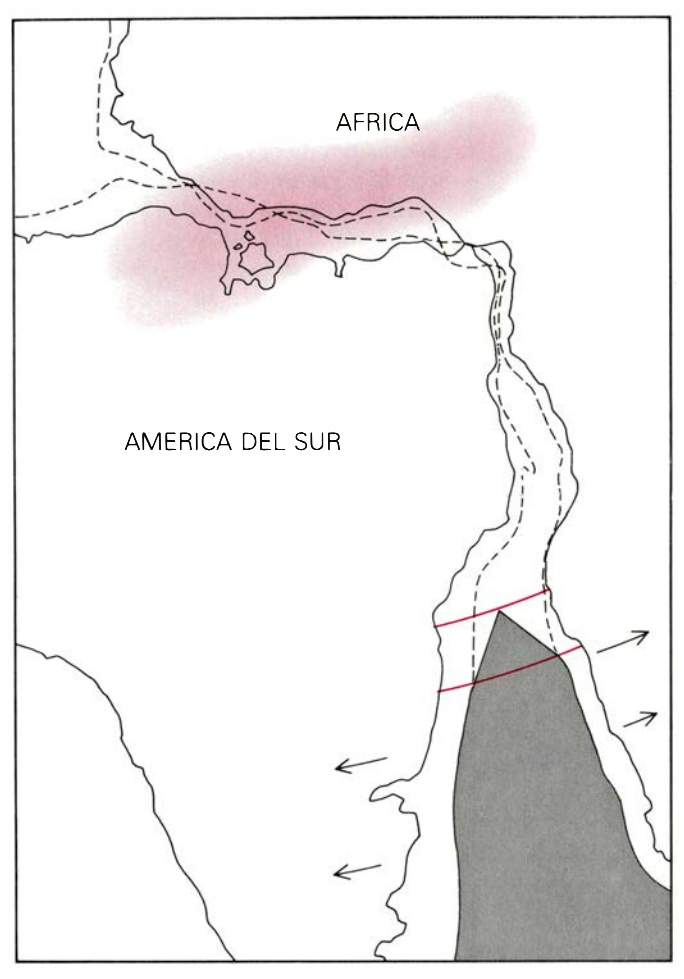
14. EXPANSION OSCILATORIA propuesta para justificar la edad anómala de la caliza extraída en Vema. En un principio la caliza (*punto rojo*) se hallaba al norte de la falla transformante de Vema y, así, se desplazó hacia el este integrada en la placa africana (*a*). Al alterarse la posición de la transformante, la caliza quedó al sur de ella (*b*). Hoy la caliza se desplaza hacia el oeste integrada en la placa americana (*c*). Un valle fósil indica la posición originaria de la transformante. (Los dibujos son obra de Andrew Tomko.)

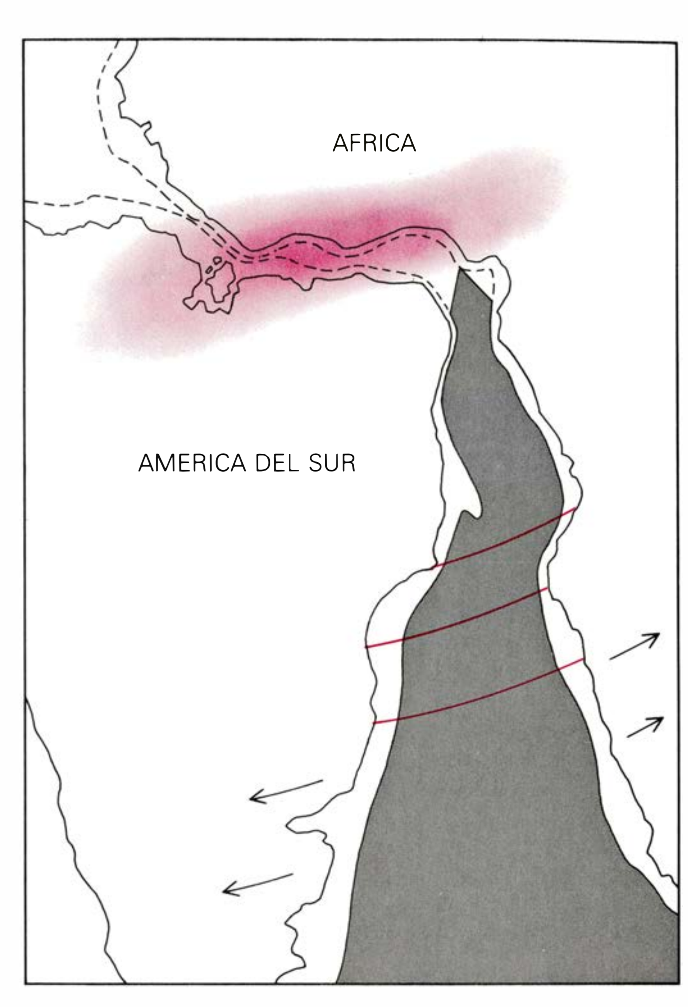
dorsal mesoceánica no pueden concebirse como accidentes inertes del fondo oceánico. Antes bien, se trata de accidentes muy activos: están sujetos a migración, a cambios de orientación y a deformación. Veamos por qué. La extensión de una falla transformante a lo largo de varios centenares de kilómetros supone que atraviesa corteza oceánica relativamente fría al pasar de un segmento de eje de dorsal a otro. Esa corteza se comporta de manera bastante rígida. Si la dislocación fuera más corta, la corteza que atraviesa estaría más caliente y sería más plástica. Los reajustes resultarían más fáciles. Tal vez no deba sorprender que tanto los accidentes topográficos como la sismicidad aumenten con la longitud del desencajamiento entre dos segmentos de ejes de dorsal.

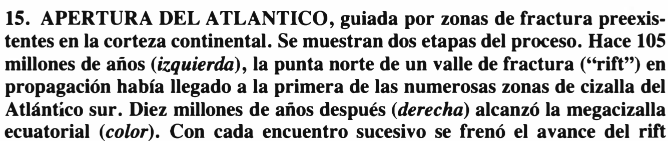
Las transformantes que cortan la dorsal mesoatlántica suelen ser mucho más accidentadas que las que segmentan la dorsal del Pacífico oriental, el gran eje de dorsal del Pacífico. También ello tiene explicación. La expansión del fondo marino es hasta 10 veces más rápida en la dorsal del Pacífico oriental que en el Atlántico. Por consiguiente, para longitudes de dislocación iguales, la corteza afectada por una transformante será más joven, más caliente y menos rígida en el Pacífico que en el Atlántico.

En última instancia, la inestabilidad de las transformantes de eje de dorsal responderá, por supuesto, a la distribución de esfuerzos en las placas oceánicas, esfuerzos que se generan por los movimientos a gran escala de las placas. Por tanto, la inestabilidad debe ser extrema cuando se fragmenta un continente y se abre una cuenca oceánica. Después de todo, la litosfera continental es más gruesa, más fría y más rígida que la litosfera oceánica. La apertura del Atlántico constituye un ejemplo que hace al caso.

El análisis del modelo de anomalías magnéticas del fondo del Atlántico actual revela que esa cuenca se abrió rompiéndose por incrementos discretos. Además, un análisis de la geometría de la cuenca induce a pensar que, tanto en el norte como en el sur del Atlántico, las líneas iniciales de apertura las determinó la posición de sistemas de fractura preexistentes que atravesaban el continente a punto de fragmentarse. Probablemente esas zonas constituían las heridas de acontecimientos tectónicos (choques entre placas, por ejemplo) anteriores a la génesis del Atlántico. De sur a norte, los grandes sistemas de fractura eran: la zona de ci-





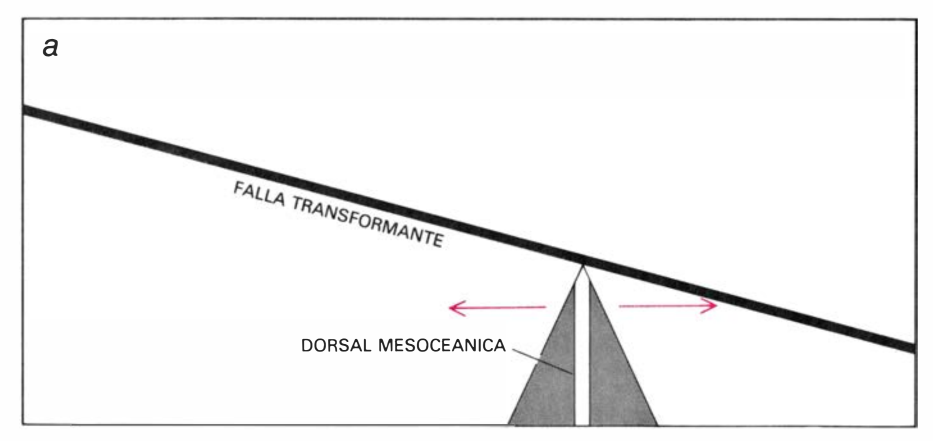

15. APERTURA DEL ATLANTICO, guiada por zonas de fractura preexistentes en la corteza continental. Se muestran dos etapas del proceso. Hace 105 millones de años (*izquierda*), la punta norte de un valle de fractura ("rift") en propagación había llegado a la primera de las numerosas zonas de cizalla del Atlántico sur. Diez millones de años después (*derecha*) alcanzó la megacizalla ecuatorial (*color*). Con cada encuentro sucesivo se frenó el avance del rift hacia el norte, hasta que del rift se expandió corteza oceánica de nuevo cuño en volumen suficiente para permitir que zona de fractura y rift se reorientaran en una dirección perpendicular a la de expansión. Detalles hipotéticos de la reorientación se representan en la figura 16. Las antiguas posiciones de Africa y América del Sur se deducen de la reubicación de sus contornos actuales (*líneas continuas*) y de la isóbata oceánica de 1000 metros (*líneas a trazos*).

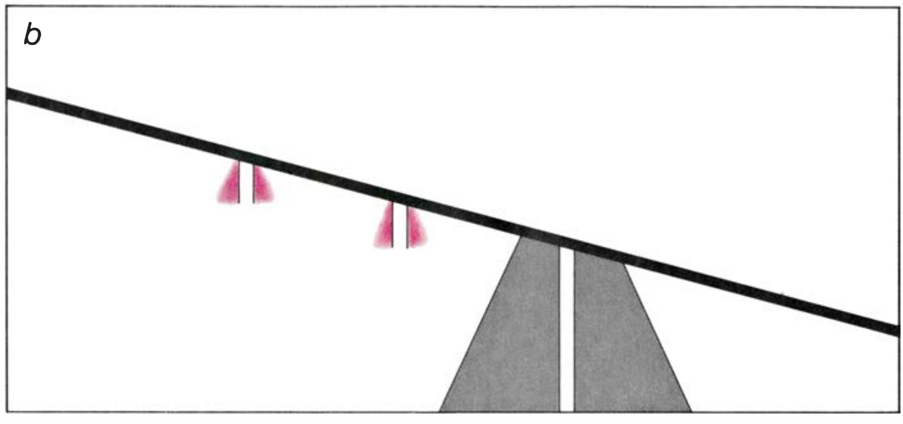
zalla de las Malvinas y la zona de megacizalla ecuatorial, dominada por la zona de fractura de Romanche: más allá, en el Atlántico norte, la zona de cizalla del Atlas y la zona de cizalla de Charlie Gibbs; y finalmente, en el mar de Noruega y Groenlandia, las zonas de cizalla de Jan Mayen y de De Geer. La evidencia que sugiere la presencia de la zona de megacizalla ecuatorial es particularmente valiosa. Los vestigios de la zona pueden identificarse en el surco amazónico, del lado americano, y en el surco de Benue, del lado africano.

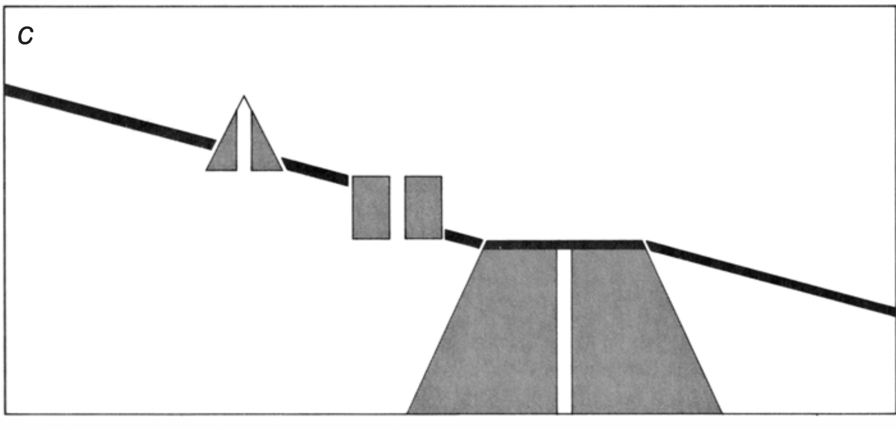
Así pues, el Atlántico sur se abrió de sur a norte, como una cremallera, a una velocidad de entre 10 y 20 centímetros por año, según los estudios de Philip D. Rabinowitz y John L. LaBreque, de Lamont-Doherty. En esencia, el extremo de un eje de dorsal mesoceánica se movió hacia el norte, dando origen a la corteza oceánica que se expandió a cada lado. El movimiento empezó en algún punto de la zona de cizalla de las Malvinas, hace unos 130 millones de años.

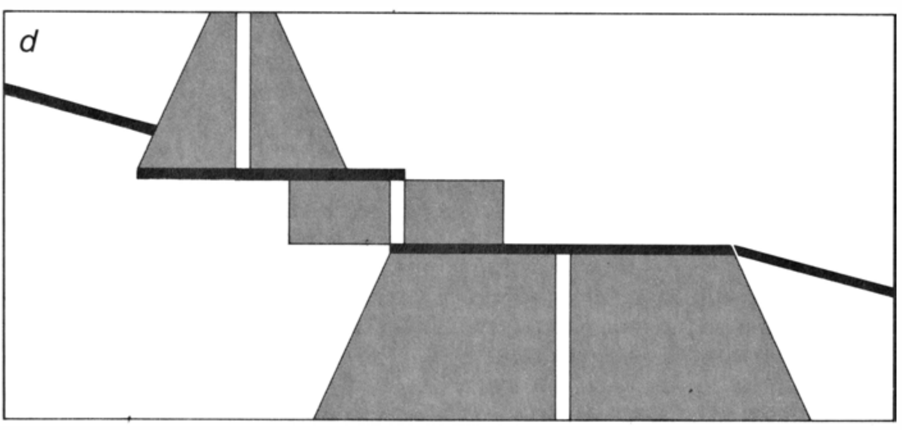
Hace 90 o 100 millones de años, el extremo del eje había llegado a la zona de megacizalla ecuatorial [*véase la figura 15*]. El encuentro debió resultar espectacular. Por una parte, la megacizalla ecuatorial bien podría representar lo que Courtillot llama una zona trabada, es decir, una parte de una placa sumamente remisa a abrirse por un valle de fractura ("rift") en propagación. Además, también es probable que el extremo del eje de propagación se cortara con la megacizalla según un ángulo no ortogonal. La megacizalla y el valle de fractura surgieron en episodios tectónicos diferentes, por lo que no hay razón para suponer que fueran mutuamente perpendiculares. La configuración oblicua habría resultado muy inestable. En un lado del eje de la dorsal, corteza oceánica recién formada se habría embutido en la megacizalla, produciendo compresión intensa y probable levantamiento. Al otro lado del eje habría ocurrido lo contrario: el esfuerzo habría sido de extensión. Tal vez la emanación de magma basáltico en la corteza debilitada creara una serie de centros locales de expansión oceánica.

Entre tanto, hacia el sur, hacia el ecuador, parecía desplazarse el extremo de un rift del Atlántico. La distancia lateral que le separaba del valle de fractura austral habría sido tan grande –más de 2000 kilómetros– que, para que se abriera el Atlántico en el ecuador, debían desarrollarse una serie de transformantes de dislocación larga. Durante no menos de 20 millones de años, esas transformantes se ubicarían en litosfera continental. Además, probablemente no fueron pependiculares a los rifts. Se crearían esfuerzos intensos, que persistirían hasta que la cantidad de corteza oceánica de nuevo cuño permitiera la conexión de los segmentos del valle y la zona de fractura para un reajuste a una configuración más perpendicular, la que encajara con la dirección de movimiento relativo de las









16. SUCESION HIPOTETICA que ilustra la incidencia de la propagación de un rift sobre una zona de cizalla preexistente. La ilustración presupone que el valle de fractura que avanza, como el que abrió el Atlántico sur, tendería a juntarse con una zona de cizalla preexistente según un ángulo oblicuo (*a*). Por tanto, en un lado del rift, la corteza acabada de formar se comprimiría contra la cizalla; en el otro lado, la corteza se estiraría y debilitaría. La corteza debilitada admitiría magma basáltico, con lo que surgirían nuevos centros de expansión oceánica (*b*). Finalmente, llegarían a producir corteza oceánica en volumen suficiente para reubicar la zona de fractura (*c, d*). En esta ilustración, uno de los nuevos centros de expansión prosigue la propagación, de manera que el eje de dorsal resulta afectado por desencajamientos.

placas. Apoyándonos en el trabajo de Dennis E. Hayes, de Lamont-Doherty, suponemos que la nueva cuenca oceánica alcanzaría por lo menos 1,5 veces la anchura de la dislocación atravesada por la zona de fractura, de modo que pudiera reajustarse toda la megacizalla ecuatorial. Sólo entonces la zona debió emigrar y girar con poco o nulo impedimento por parte de litosfera continental.

Sin duda en la apertura del Atlántico ecuatorial ha influido de forma notable la presencia de esa megacizalla ecuatorial. Se explica así que el eje de la dorsal mesoatlántica presente la impresionante serie de largas dislocaciones del ecuador. También la configuración global del Atlántico da testimonio de ello. Los valles de fractura que abrieron el océano tendieron a separar Europa y Africa de las Américas a lo largo de una línea aproximadamente norte-sur. En cambio, en el ecuador, la zona de megacizalla orientó la línea de este a oeste. Pueden verse trazas de la separación inicial en los desencajamientos escalonados, en sentido este-oeste, de las costas ecuatoriales atlánticas, tanto en el continente africano como en el sudamericano.

En la ulterior evolución del Atlántico, las transformantes ecuatoriales han mantenido su actividad, y así todos los cambios en la dirección de expansión del fondo marino arrastraron consecuencias espectaculares. En cada caso, el reajuste de la geometría de las fallas transformantes ecuatoriales habrá llevado a episodios de extensión y compresión, generando vulcanismo y movimiento vertical de astillas corticales. En uno de los levantamientos resultantes a lo largo de la transformante de San Pablo, las aves observadas por Darwin construían sus nidos en el manto terrestre. En otras partes se han hundido arrecifes coralinos fósiles en las profundidades del mar. Es probable que en el joven y angosto Atlántico ecuatorial se produjera a menudo el ascenso y descenso de bloques corticales. De este modo, la geología de las zonas de fractura oceánicas probablemente influyera en acontecimientos tan diversos como las migraciones de organismos entre continentes y los intercambios de agua entre el océano septentrional y el meridional. La intuición poética de William Carlos Williams fue penetrante cuando escribió:

Oh poderosísimo nexo,
perla entre continentes
que un hilo atraviesa...

Endocitosis de proteínas y partículas

La endocitosis mediada por receptor es un proceso por el que las células captan macromoléculas específicas. En la mayoría de los casos, el receptor, después de liberar su ligando, vuelve a la membrana plasmática para capturar más ligando

Alice Dautry-Varsat y Harvey F. Lodish

Las células de los organismos pluricelulares se hallan rodeadas de un medio acuoso, derivado de la sangre, que viene a ser como un "caldo enriquecido". Se trata de un caldo poco usual, por cuanto contiene varios miles de ingredientes, la mayoría de los cuales se presentan en concentraciones sumamente bajas. Algunos de los ingredientes son materiales estructurales de la célula, como los aminoácidos, y nutrientes, como las vitaminas, que las células necesitan en cantidades y en momentos precisos. Otros son hormonas, que transmiten señales intracelulares específicas. También hay productos de desecho, e incluso sustancias tóxicas, que degradan células especiales equipadas al efecto. Cada célula debe procurarse del medio extracelular las sustancias que necesita, cerrando el paso a las que no le convienen.

La membrana plasmática es el tamiz que delimita la célula y controla su tráfico de productos con el medio y, por tanto, con las demás células del organismo. Como todas las membranas biológicas, la plasmática es, en esencia, una doble capa de moléculas fosfolipídicas en las que están embebidas gran variedad de moléculas de proteínas. Las proteínas desempeñan un amplio abanico de funciones, una de las cuales es facilitar la toma selectiva de sustancias específicas solubles en agua a través de la bicapa lipídica, por lo demás impermeable. Los iones (átomos cargados) y moléculas pequeñas solubles en agua, como los aminoácidos (los constituyentes de las proteínas) y los azúcares, fluyen por canales especializados de la membrana o se bombean a través de los mismos; los canales están compuestos por las proteínas llamadas permeasas. Las moléculas de gran tamaño y la materia particulada se introducen en la célula por un proceso totalmente diferente, llamado endocitosis: una porción de la membrana plasmática envuelve el material que debe incorporarse; así se introduce en la célula, encerrado en una vesícula derivada de la membrana plasmática. Se distinguen tres clases de endocitosis.

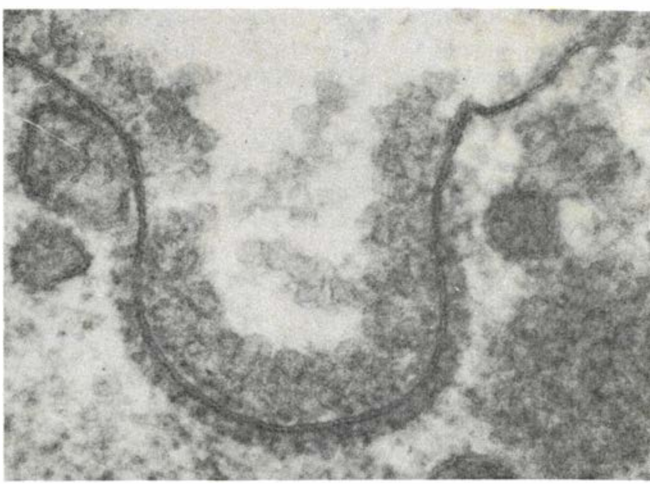
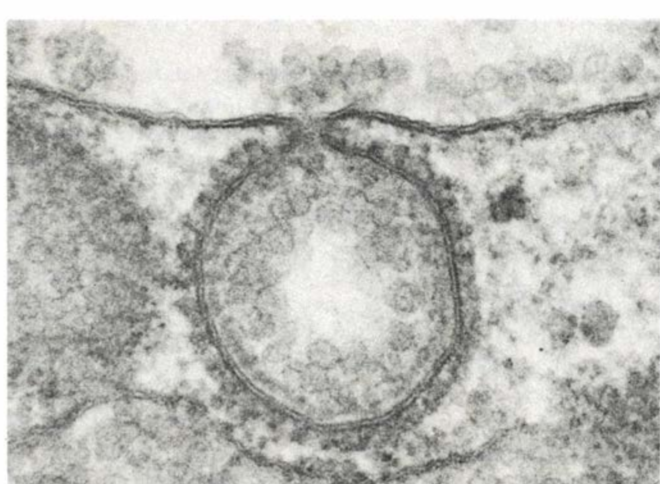
En la fagocitosis, la unión a la superficie celular de una partícula muy grande, o de un complejo molecular, provoca la expansión de la membrana alrededor del objeto, que se incorpora a la célula en una vesícula sacular, una porción invaginada de la membrana que puede alcanzar varios micrometros de diámetro. La fagocitosis es el proceso por el que los protozoos ingieren bacterias y otras partículas alimenticias; en los animales superiores, la depredación por fagocitosis de elementos invasores, como las bacterias, corresponde a ciertas células del sistema inmunitario denominadas macrófagos.

La pinocitosis, un proceso diferente, consiste en la toma inespecífica de fluido extracelular. Una diminuta gotita de líquido queda envuelta por una porción de membrana plasmática invaginada y se introduce en una vesícula que no mide más de 0,1 micrometros de diámetro; se incorporan así cualesquiera iones o moléculas pequeñas que puedan hallarse en la gota.

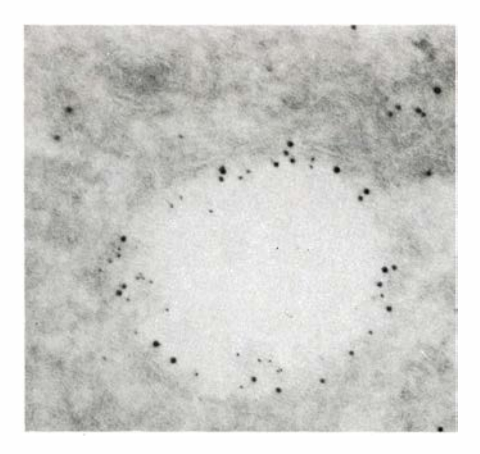
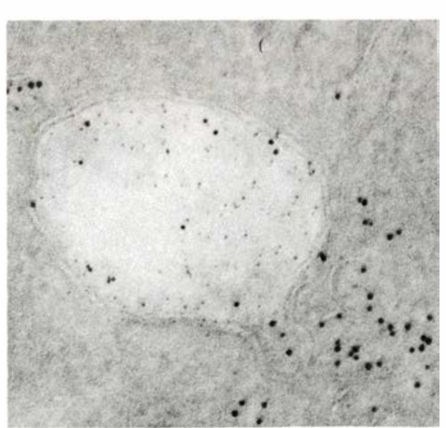
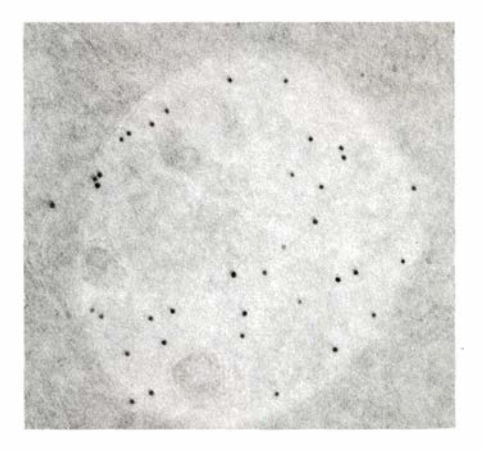
La endocitosis mediada por receptor resulta, en contraposición, exquisitamente específica. Los receptores son proteínas de membrana, cada uno de los cuales tiene un centro de unión en el que encaja un determinado ligando: una proteína o una partícula pequeña. En efecto, el receptor toma un ingrediente del "caldo" extracelular (por muy baja que sea su concentración externa y exista un gran exceso de moléculas no relacionadas con ella) y lo ase fuertemente. La unión provoca una invaginación de la membrana, formándose una vesícula unida a la membrana que encierra el ligando y lo introduce en la célula. En los últimos años se ha avanzado notablemente en la comprensión del mecanismo de la endocitosis mediada por receptor; en particular, de los principales procesos de la disociación del complejo receptor-ligando internado, el envío del ligando a su destino intracelular y el reciclaje del receptor hasta la superficie de la célula para atrapar más ligando.

Una amplia variedad de ligandos se incorporan a las células mediante la endocitosis mediada por receptor; una vez dentro de ellas, se emplean de diferente forma. El primer ejemplo de la toma de proteína por este método lo describieron, en 1964, Thomas F. Roth y Keith R. Porter, a la sazón en la Universidad de Harvard. Se explicaba así cómo obtenían los huevos de aves e insectos la gran cantidad de proteínas almacenada en la yema para nutrir el embrión: las proteínas se importaban al interior del ovocito en desarrollo después de haberse sintetizado en otro lugar. Roth y Porter demostraron que, en el mosquito, la proteína precursora vitelogenina se sintetizaba en el hígado de la hembra, se secretaba a la sangre y se transportaba al ovario, donde se unía a receptores en unos 300.000 hoyuelos de la superficie del ovocito. Los hoyos penetraban en forma de vesículas, que se fundían para constituir la yema. Enzimas digestivos de proteínas (proteasas) escindían la vitelogenina en dos proteínas esenciales de la yema, lipovitelina y fosvitina.

En los mamíferos, la inmunidad materna se transfiere al feto en desarrollo por endocitosis mediada por receptor. Los anticuerpos de la sangre de la madre se unen a las células fetales que cubren el saco vitelino; está dotado éste


1. ETAPAS SUCESIVAS DE LA ENDOCITOSIS, según una serie de micrografías electrónicas de ovocitos de gallina, precursores de los óvulos. La suave endentación de la membrana plasmática del ovocito (*arriba, izquierda*) es un hoyo revestido. El revestimiento grueso sobre la cara citoplasmática (interna) del hoyo es una malla de la proteína clatrina. Se supone que las grandes partículas que revisten la membrana por la cara extracelular del hoyo son partículas lipoproteicas; probablemente estén unidas a receptores específicos de la membrana del ovocito. En etapas posteriores, el hoyo se hace más profundo (*arriba, derecha*), se redondea para formar un esbozo de vesícula revestida (*abajo, izquierda*) y la vesícula se separa de la superficie de la membrana, introduciendo en la célula el ligando de lipoproteína (*abajo, derecha*). Las micrografías son de M. M. Perry y A. B. Gilbert, del Consejo de Investigación Agrícola de Edimburgo. Trataron las células con ácido tánico y las tiñeron con citrato de plomo. La ampliación es de 135.000 diámetros.

2. SEPARACION DEL RECEPTOR Y EL LIGANDO en células de hígado de rata, visualizada por una técnica de microscopía electrónica de doble marcaje. Las células, congeladas y seccionadas, se trataron secuencialmente con una preparación específica de anticuerpos para el ligando (una proteína con galactosa terminal anormal) y con otra específica para su receptor; los dos anticuerpos se marcaron indirectamente con partículas de oro de diferente tamaño. Los anticuerpos se unieron a sus antígenos respectivos, cuya localización revelaron las partículas de oro densas a los electrones, apreciables en las micrografías como puntos grandes y pequeños. En un endosoma (*izquierda*), una vesícula de endocitosis próxima a la superficie celular, las moléculas de receptores (*puntos grandes*) y de lingado (*puntos pequeños*) aparecen juntas en la superficie interna de la membrana vesicular. A mayor profundidad, en una vesícula llamada CURL (*centro*), las moléculas de ligando se han liberado de los receptores y están libres en el interior de la vesícula; los receptores se concentran en estructuras tubulares unidas al lado inferior derecho de la vesícula. Se supone que esos túbulos membranosos devuelven los receptores a la superficie celular. Una micrografía tomada con marcaje de oro invertido muestra una vesícula en etapa avanzada, un cuerpo multivesicular (*derecha*): contiene moléculas de ligando pero no de receptor. Las micrografías las realizaron Hans J. Geuze, Jan W. Slot y Ger J. A. M. Strous, de la Universidad de Utrecht, en colaboración con Alan L. Schwartz y uno de los autores (Lodish). Las secciones aparecen aquí ampliadas unos 90.000 diámetros.

de receptores superficiales que reconocen una región común a todos los anticuerpos circulantes del tipo inmunoglobulina G. Los anticuerpos se transportan al interior de las células y se secretan a la circulación fetal.

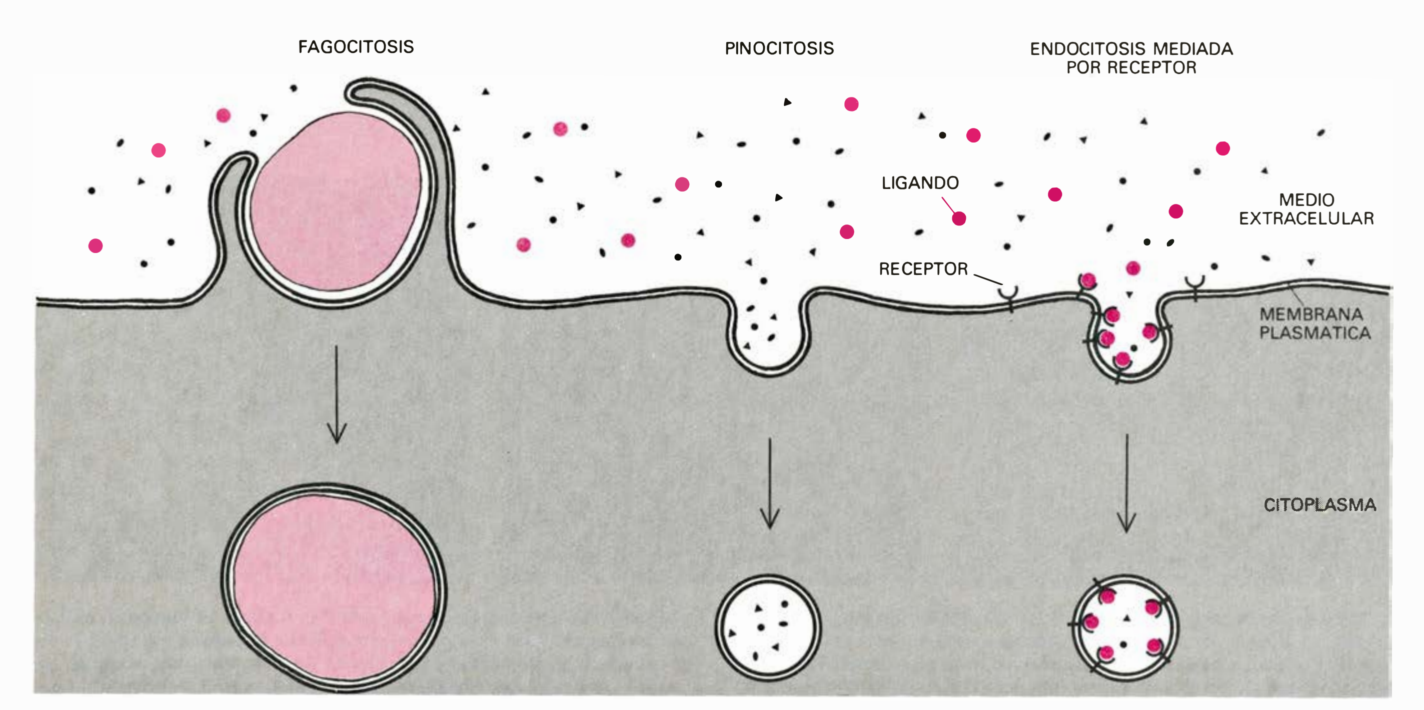
El colesterol es un componente lipídico mayoritario de la membrana plasmática de todas las células de mamíferos. Richard G. W. Anderson, Michael S. Brown y Joseph L. Goldstein, del Hospital Clínico de la Universidad de Texas en Dallas, han demostrado con elegante precisión cómo obtiene colesterol la célula. Este se sintetiza y almacena en el hígado; allí se empaqueta en grandes partículas esféricas lipoproteicas de baja densidad (LDL, por *low-density lipoprotein*), que hacen las veces de transportador. Cada partícula LDL tiene un núcleo de unas 1500 moléculas de colesterol, unidas químicamente a cadenas de ácidos grasos. El núcleo está envuelto en una membrana fosfolipídica de una sola capa y otras moléculas de colesterol; la membrana contiene también una proteína de unión, llamada *apo-b*. Del hígado, la partícula se secreta a la circulación.

Las partículas LDL se unen a receptores específicos de la superficie celular que reconocen la *apo-b*; penetran en la célula por vesículas unidas a la membrana y se transportan, en una serie de vesículas progresivamente mayores, al lisosoma: un orgánulo celular equipado con hidrolasas, es decir, enzimas digestivas. Las enzimas escinden las partículas LDL, degradan la proteína de unión, separan los ácidos grasos de las moléculas de colesterol y liberan el colesterol y los ácidos grasos para su incorporación en la membrana celular. Brown, con Goldstein y sus colaboradores, han probado que la falta de funcionalidad de los receptores de LDL es responsable de la hipercolesterolemia familiar, enfermedad genética caracterizada por la presencia de niveles extremadamente altos de colesterol en sangre, y la consecuente aterosclerosis precoz.

El hierro es un constituyente esencial de todas las células, a las que llega por medio de una proteína transportadora llamada transferrina. El transportador toma iones férricos (Fe^{+++}) en el intestino (donde el hierro se absorbe de los alimentos) y en el hígado (donde se almacena). La molécula de ferrotransferrina cargada con dos iones de hierro se une a un receptor específico de la superficie celular, y el complejo receptor-transferrina se introduce por endocitosis; se aporta, así, hierro a la célula.

Una de las funciones más importantes de la endocitosis mediada por receptor es introducir hormonas y otras proteínas que liberan señales específicas a ciertas células. Las moléculas de insulina, por ejemplo, se unen a receptores específicos de las células diana. La unión desencadena varios procesos metabólicos; entre otros, un aumento de la toma de glucosa. Se cree que la respuesta hormonal termina cuando se introducen los complejos insulina-receptor y la insulina se degrada en los lisosomas. Los materiales potencialmente perjudiciales se desechan también de esa forma. Por ejemplo, ciertas glicoproteínas (proteínas con cadenas de azúcares unidas a ellas) anormales tienen azúcar galactosa en el extremo de la cadena, en vez del azúcar usual, ácido siálico. Las células del hígado tienen receptores que reconocen y se unen a las glicoproteínas con galactosa terminal, las introducen en la célula y las degradan en los lisosomas.

Los receptores que median la endocitosis son proteínas que atraviesan de un lado a otro la membrana plasmática. Son anfipáticos, es decir, tienen dos regiones hidrofílicas, que se proyectan al medio acuoso exterior de la célula y al citoplasma, y una región central hidrofóbica, que se une fuertemente a los ácidos grasos que constituyen el armazón de la membrana. No se conoce en detalle la estructura de tales receptores, pero se sabe que el receptor de la transferrina es una glicoproteína con un peso molecular de 180.000. Ian S. Trowbridge, del Instituto Salk de Estudios Biológicos, Robert Allen, de la Universidad de Colorado,



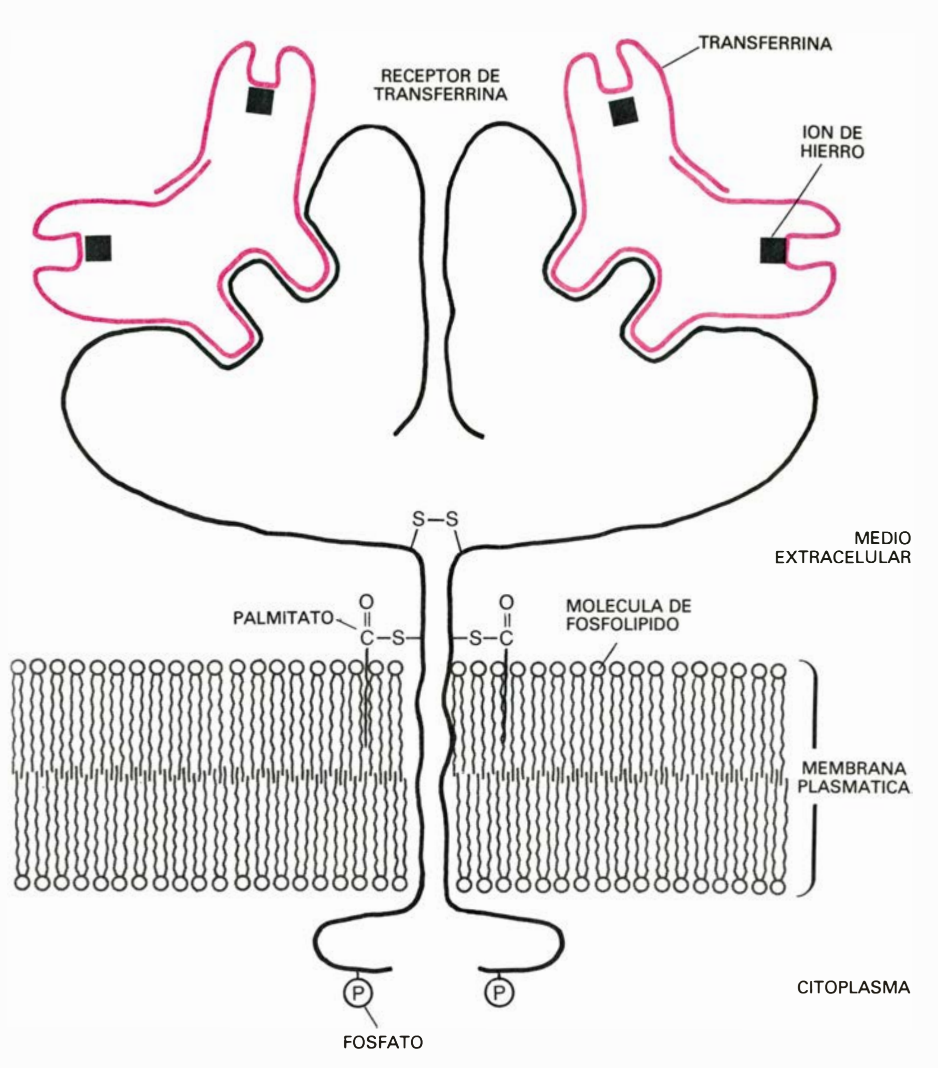
3. TRES CLASES DE ENDOCITOSIS, en esquema. Comencemos por la fagocitosis (*izquierda*): la membrana plasmática de la célula se une a una partícula grande, como pueda ser una bacteria, y se expande envolviéndola y engulléndola hacia el interior de la célula. En la pinocitosis (*centro*), una porción de la membrana rodea una gotita de medio líquido extracelular; la membrana se pliega hacia el interior, formando una vesícula que encierra la gotita y cualesquiera pequeñas moléculas disueltas en ella. La endocitosis mediada por receptor (*derecha*) es un mecanismo para la toma selectiva de partículas o macromoléculas. Un ligando se une a su receptor específico sobre la membrana plasmática, provocando la introducción del complejo receptor-ligando en una invaginación de la membrana plasmática. La vesícula así formada se desprende hacia el interior de la célula, transportando con ella el ligando.

y Howard H. Sussman, de la Universidad de Stanford, han demostrado que el receptor de transferrina consta de dos cadenas polipeptídicas idénticas, cada una de ellas de una longitud de aproximadamente 800 aminoácidos, enlazadas por un único enlace disulfuro. Cada cadena polipeptídica tiene, por lo menos, tres cadenas de carbohidratos, además de un ácido graso, palmitato, que quizás ayude al anclaje del receptor en la membrana.

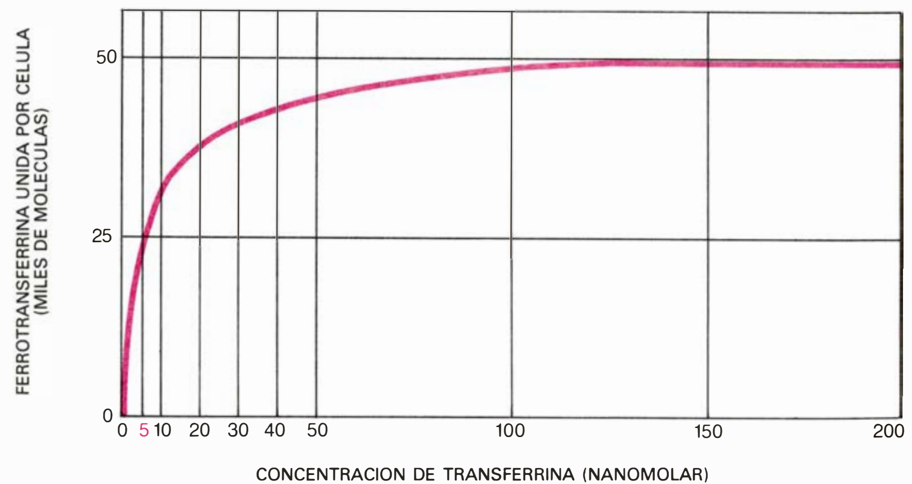
Cada receptor se une a dos moléculas de transferrina, probablemente una molécula por cadena polipeptídica. En condiciones fisiológicas, a la unión sigue rápidamente la incorporación. Cabe separar los dos pasos con propósitos experimentales, ya que la unión no requiere energía, mientras que la introducción sí la requiere; ello permite bloquear ese paso por temperaturas bajas o por inhibidores de la producción celular de energía. En tales condiciones, la unión se produce normalmente y se puede cuantificar. Como la mayoría de los receptores, el de transferrina muestra una afinidad muy alta para su ligando. Su constante de disociación (la concentración de ligando a la cual la mitad de los receptores están ocupados) es 5 nanomolar (5×10^{-9} M), es decir, unos 350 microgramos por litro. Se sigue de ello que la transferrina permanece enlazada a su receptor incluso cuando la concentración de transferrina es sólo una cienmilésima de la concentración total de proteínas en la sangre.

Los complejos ligando-receptor se agrupan en ciertos sitios de la membrana plasmática, concretamente en los llamados hoyos revestidos (*coated pits*). Se ha demostrado que los hoyos, descubiertos por Roth y Porter, se presentan en la mayoría de las células animales; típicamente ocupan un 2 por ciento de la superficie celular. El revestimiento del que los hoyos toman el nombre, una gruesa capa proteica ubicada en el lado interno de la membrana plasmática, por debajo de cada hoyo, está constituido principalmente por clatrina, una proteína fibrosa que identificara Barbara M. F. Pearse, del Laboratorio de Biología Molecular del Consejo de Investigación Médica de Inglaterra.

La región de hoyos revestidos de la membrana es la que se invagina para formar una vesícula. Se trata de un proceso continuo. Los hoyos revestidos se mantienen plegados hacia el interior para formar vesículas y se regeneran continuamente en la superficie de la célula. Las propias vesículas están revestidas, pero ahora, por supuesto, sobre su superficie externa o citoplasmática:



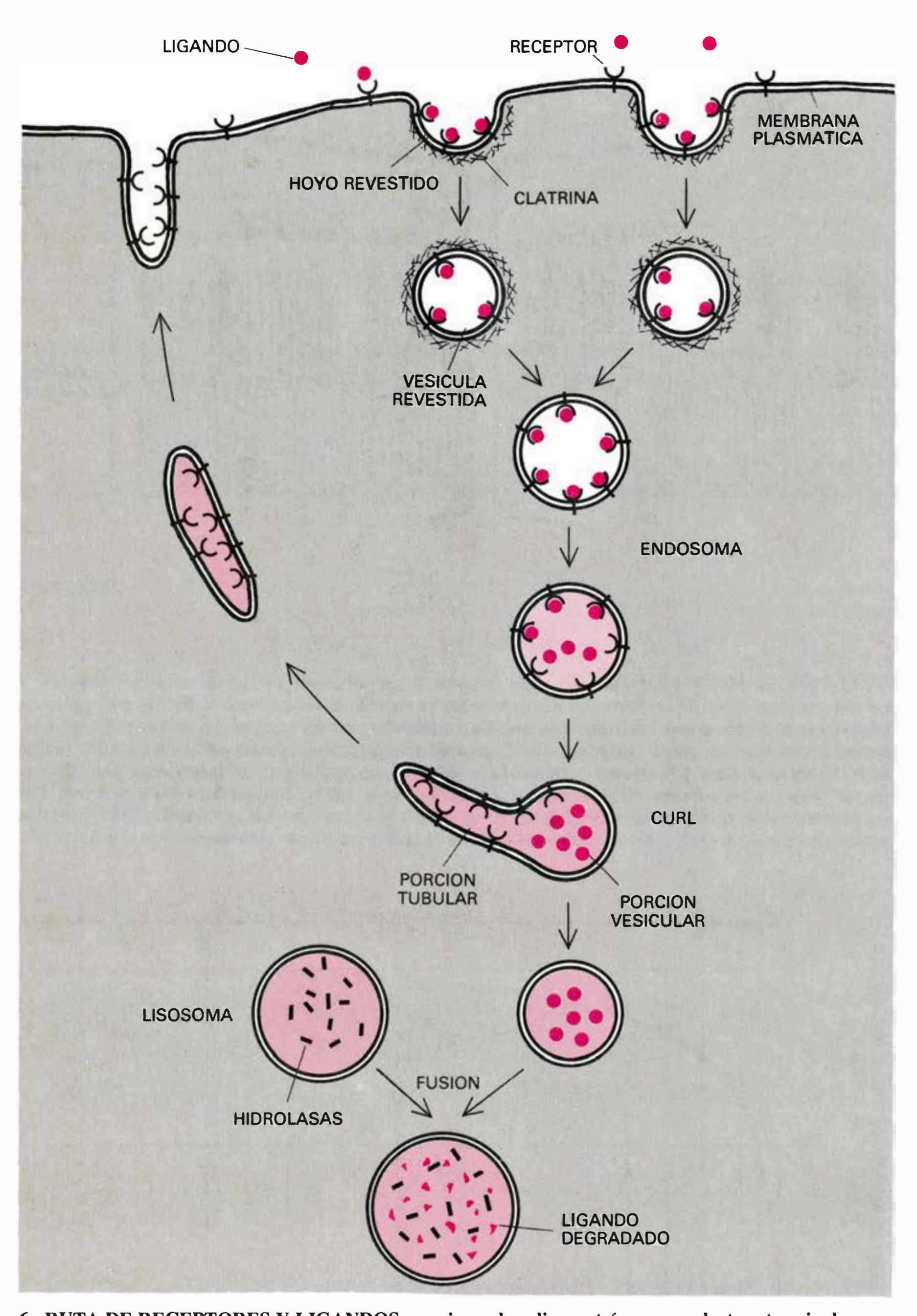
4. RECEPTOR DE TRANSFERRINA, una proteína transportadora de hierro; está embebido en la membrana plasmática de la célula. La membrana es, en esencia, una bicapa de moléculas fosfolipídicas, cuyas colas de ácidos grasos la hacen impermeable a moléculas solubles en agua. No se conoce la estructura real del receptor de transferrina; en este dibujo, muy esquemático, su forma refleja lo poco que se sabe de él. El receptor tiene dos cadenas polipeptídicas idénticas, enlazadas por un puente disulfuro (*S-S*); la mayor parte de las cadenas están expuestas al exterior de la célula; cada cadena tiene un ácido graso, palmitato, que quizás ayude al anclaje del receptor; éste une dos moléculas de transferrina, probablemente una molécula por cadena. Cada molécula de transferrina puede transportar dos iones férricos.



5. CALCULO DE AFINIDAD de los receptores a la transferrina cargada de hierro. Se incubaron células de tumor hepático en concentraciones crecientes de transferrina marcada. El ligando se une a la superficie, pero no penetra en el interior celular; se expresa aquí la proporción cuantitativa de enlace. El nivel superior de enlace, 50.000 moléculas por célula, demuestra que las células poseen unos 50.000 receptores de transferrina sobre su superficie. La concentración a la que se encuentran ocupados la mitad de ellos, 5 nanomolar, es su constante de disociación, un medidor de la afinidad del receptor por su ligando.

el revestimiento de clatrina envuelve la membrana celular en una red fibrosa de pentágonos y hexágonos. (Por lo que se ve, la polimerización de clatrina forma en realidad la vesícula revestida. Si se purifican las vesículas revestidas hasta liberarlas de membrana, puede disociarse la clatrina en estructuras de subunidades de tres brazos llamadas trisqueliones. En condiciones apropiadas, los trisqueliones se agregan componiendo el tipo de estructura en forma de cesto que rodea las vesículas revestidas.) Al adentrarse más profundamente en el citoplasma, las vesículas revestidas se desprenden de su revestimiento de clatrina y se funden entre sí (o con alguna otra clase de vesículas) en vesículas mayores, de superficie lisa, llamadas endosomas.

¿Residen siempre los receptores en hoyos revestidos, o acaso la unión de un ligando les impulsa a emigrar a través de la membrana, que constituye un sustrato esencialmente fluido? Depende del receptor. Anderson ha demostrado que los receptores de LDL de fibroblastos (células de tejido conectivo) se hallan principalmente en hoyos revestidos, aun cuando no estén unidos a partículas LDL. Por el contrario, los receptores de transferrina, insulina y glicoproteínas con galactosa terminal suelen presentarse dispersos por la membrana plasmática; se agrupan en hoyos revestidos cuando se enlaza el ligando, y siempre que la temperatura no sea inferior a los 37 grados Celsius. Una explicación sencilla a esa discrepancia sería que ciertos receptores gozan de determinadas propiedades que los dirigen hacia los hoyos revestidos y los mantienen en ellos; parte del receptor podría, por ejemplo, mostrar afinidad para un componente de la región del hoyo. Tal propiedad podría aparecer en otros receptores sólo tras la unión a su ligando, quizás a través de un cambio de la conformación del receptor.



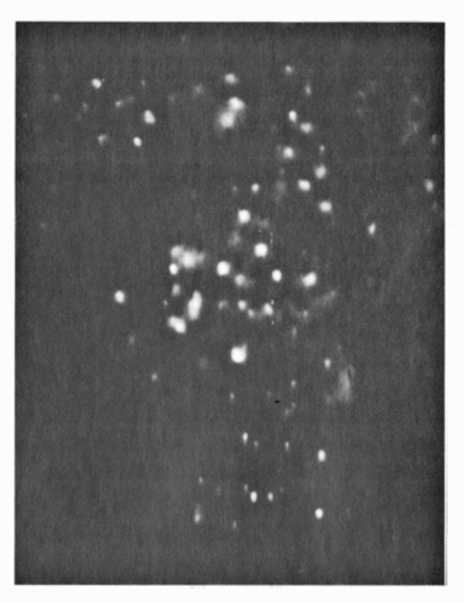
6. RUTA DE RECEPTORES Y LIGANDOS que siguen las glicoproteínas con galactosa terminal, y que probablemente también sea aplicable a otros lingados y receptores. El ligando se une a sus receptores, distribuidos difusamente por la membrana y que entonces se agrupan en los hoyos revestidos. Estos se invaginan e introducen como vesículas revestidas, cuya fusión produce endosomas, y luego vesículas CURL (por *compartment of uncoupling of receptor and ligand*, compartimento de desenganche del receptor y el ligando). En el medio ácido de la vesícula CURL (*color*), el ligando se disocia de los receptores, se acumula en el lumen de la vesícula CURL y los receptores se concentran en la membrana de una estructura tubular adosada, que se separa entonces de la vesícula CURL. La porción vesicular se mueve hacia el interior celular y se funde con un lisosoma, al que libera el ligando para su degradación. Se cree que la estructura tubular membranosa torna los receptores a la membrana plasmática. Los dibujos son de Bunji Tagawa.

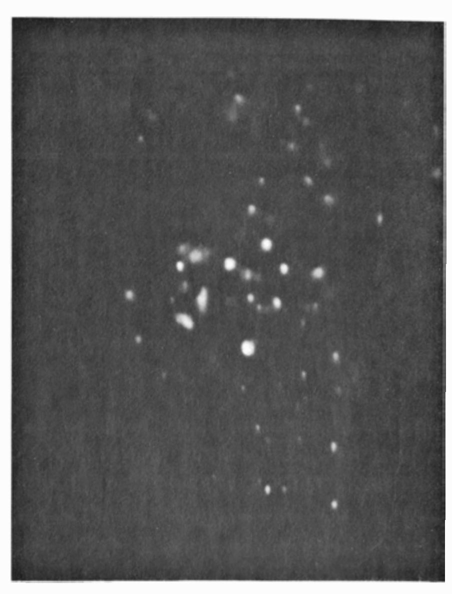
Aunque cada ligando tiene su propio destino, la mayoría de ellos acaban en los lisosomas, donde unas 40 enzimas digestivas que operan en medio ácido (a un *p*H de 4,5 a 5, en contraste con el *p*H neutro, de 7 a 7,4, del citoplasma) degradan los ligandos. Los productos de digestión se eliminan de la célula o bien se exportan al citoplasma en calidad de materia prima (como es el caso de los componentes del LDL). Sería un derroche que los receptores compartieran el destino de sus ligandos; por supuesto, no ocurre así. Hace años se comprobó que los receptores, una vez liberados sus ligandos en el interior de la célula, volvían a la superficie celular para recomenzar el proceso. Diversas pruebas lo sugerían. Aunque se bloqueara la formación de nuevos receptores con inhibidores específicos de la síntesis de proteínas, persistía la captura de ligandos. Se demostró que la vida media de la mayoría de los receptores excedía con mucho la vida media de sus ligandos, y que los receptores de transferrina, LDL o glicoproteínas con galactosa terminal introducían un ligando cada 10 o 15 minutos, persistiendo en su función durante muchas horas.

¿Opera el reciclaje de los receptores de forma continua, como los autobuses urbanos, o depende de la inducción que pueda desencadenar la unión de un ligando, a modo de taxis? De nuevo la respuesta no es única para todos los receptores. Los de LDL parecen mantener su ciclo en los fibroblastos humanos

con independencia de que porten o no un ligando LDL. Por otra parte, hay pruebas de que los receptores de glicoproteínas con galactosa terminal se introducen sólo si llevan ligando. Aparentemente, los receptores localizados siempre en hoyos revestidos (incluso en ausencia de ligando) se reciclan continuamente, porque los hoyos revestidos se invaginan sin cesar; otros receptores se reciclan sólo en presencia de ligando, porque sólo entonces se mueven hacia los hoyos revestidos.

¿En qué etapa, y cómo, se separan los receptores de sus ligandos? Hans J. Geuze, Jan W. Slot y Ger J. A. M. Strous, de la Universidad de Utrecht, en colaboración con Alan L. Schwartz y uno de los autores (Lodish), emplearon una técnica nueva de microscopía electrónica para averiguar dónde tenía lugar la separación. Se fabricó un anticuerpo dirigido contra una glicoproteína con galactosa terminal y otro dirigido contra el receptor de galactosa; se marcaron los dos anticuerpos con partículas de oro de diferente tamaño. Se inyectó en ratas la glicoproteína y se sacrificaron aquéllas tras distintos intervalos de espera; se fijaron sus hígados y los cortes obtenidos de ellos se trataron con los dos anticuerpos, que se unían al ligando y al receptor; las partículas de oro se manifestaron en las micrografías electrónicas como puntos densos.





7. LA ACIDEZ de las vesículas de endocitosis quedó demostrada en un ensayo de Benjamin Tycko y Frederick R. Maxfield, del Hospital Clínico de la Universidad de Nueva York. Se marcó un ligando, la proteína alfa-2-macroglobulina, con fluoresceína, un colorante fluorescente sensible al *p*H. Se incubaron cultivos de fibroblastos (células de tejido conectivo) con el ligando, el tiempo justo para que el ligando alcanzara los endosomas o vesículas CURL. Entonces se iluminaron las células con luz violeta de 450 nanómetros o con luz azul de 490 nanómetros, y se observaron con una cámara de video intensificadora de imagen montada en un microscopio de fluorescencia. La relación de las intensidades de fluorescencia, bajo iluminación a las dos longitudes de onda, proporciona una medida del *p*H interno de las vesículas que contienen el ligando marcado. La fluorescencia a 450 nanómetros (*izquierda*) brilla algo más que la fluorescencia a 490 nanómetros (*derecha*), señal de que el *p*H es ácido; la calibración mediante las relaciones de intensidad medidas en disoluciones de *p*H conocido indica que el *p*H de esas vesículas es de 5,0.

En cortes practicados en las primeras etapas de la endocitosis, el receptor y su ligando aparecían estrechamente asociados con la membrana de las vesículas, disponiéndose justo debajo de la superficie de la célula. El lumen (el interior ocupado por fluido) de las vesículas no contenía ligando libre. Aparentemente, el ligando permanecía unido a los receptores, embebidos en la membrana de la vesícula. Sin embargo, más lejos de la superficie de la célula se distinguían vesículas mayores en las que sí había ligando libre en el lumen; cuanto mayores eran las vesículas, más ligando libre había. En esas vesículas más grandes, los receptores aparecían todavía asociados con la membrana, pero no se distribuían al azar por la membrana vesicular; antes bien, los receptores se agrupaban cerca de un extremo de la vesícula, donde ésta parecía adyacente o estaba realmente fusionada con túbulos membranosos delgados, o bien se hallaban en los propios túbulos.

Esa estructura de vesícula con túbulo la hemos bautizado vesícula CURL (por *compartment of uncoupling of receptor and ligand*, compartimento de desenganche de receptor y ligando). Tal parece ser el orgánulo en el que se separan y redistribuyen receptores y ligandos, acumulándose los ligandos en la parte vesicular y los receptores en los túbulos membranosos. Los cortes de tejido practicados cuando la endocitosis ha progresado más muestran que las vesículas CURL se fusionan en última instancia con lisosomas, donde se degrada el ligando. Sin embargo, antes de producirse la fusión, las porciones tubulares cargadas con receptores se separan de las vesículas CURL; los receptores escapan así a la degradación lisosómica. Las estructuras tubulares parecen servir de intermediarios que devuelven los receptores a la superficie de la célula. No se sabe aún si se trata de vesículas que se dirigen a la superficie o si forman parte de un sistema tubular persistente, análogo al retículo endoplasmático, que se extiende hacia la superficie.

En la superficie celular, los ligandos se unen a sus receptores con gran fuerza. ¿Qué provoca la disociación entre receptor y ligando en el interior de la célula? El *p*H bajo. Hace pocos años se publicó que el reciclaje de los receptores podía inhibirse por medio de agentes que aumentaban el *p*H de las vesículas ácidas del interior de la célula. Benjamin Tycko y Frederick R. Maxfield, del Hospital Clínico de la Universidad de Nueva York, probaron de forma directa que los endosomas (y probablemente también las vesículas CURL) solían ser ácidas. Estudiaron la endocitosis de la alfa-2-macroglobulina en fibroblastos humanos, una proteína que se une a proteasas y que se liga a receptores de la superficie celular que la transfieren al interior, con lo que se eliminan proteasas de la circulación. Tycko y Maxfield acoplaron colorante fluorescente fluoresceína a la alfa-2-macroglobulina. El espectro de excitación de la fluoresceína es sensible al *p*H, es decir, la intensidad de su fluorescencia varía con el *p*H, así como con la longitud de onda de la luz a la que se expone el colorante.

Se incubaron los fibroblastos con ligando marcado con fluoresceína el tiempo justo para que el ligando se incorporara y alcanzara las vesículas prelisosómicas, como endosomas o vesículas CURL. A continuación se enfriaron las células, se iluminaron a dos longitudes de onda diferentes y se observaron con una cámara intensificadora de imagen montada sobre un microscopio de fluorescencia. Los cálculos basados en la intensidad relativa de fluorescencia a dos longitudes de onda de excitación indicaron que el *p*H de las vesículas que contenían el ligando era de 5,0.

Los experimentos con receptores y ligandos purificados muestran que, a *p*H

neutro, los receptores se unen fuertemente a sus ligandos, mientras que cuando el pH baja a 5,5, los ligandos se disocian rápidamente. Está claro que cuando el complejo receptor-ligando entra en una vesícula ácida, el ligando se separa del receptor y se solubiliza en el lumen de la vesícula; el receptor permanece unido a la membrana vesicular y acaba reciclándose. No se ha establecido todavía cómo se acidifican los endosomas (o las vesículas CURL), pero se sabe de una enzima de la membrana endosómica que utiliza la energía almacenada en el adenosín trifosfato (ATP) para bombear protones hacia el lumen. Un exceso de protones significa un pH bajo.

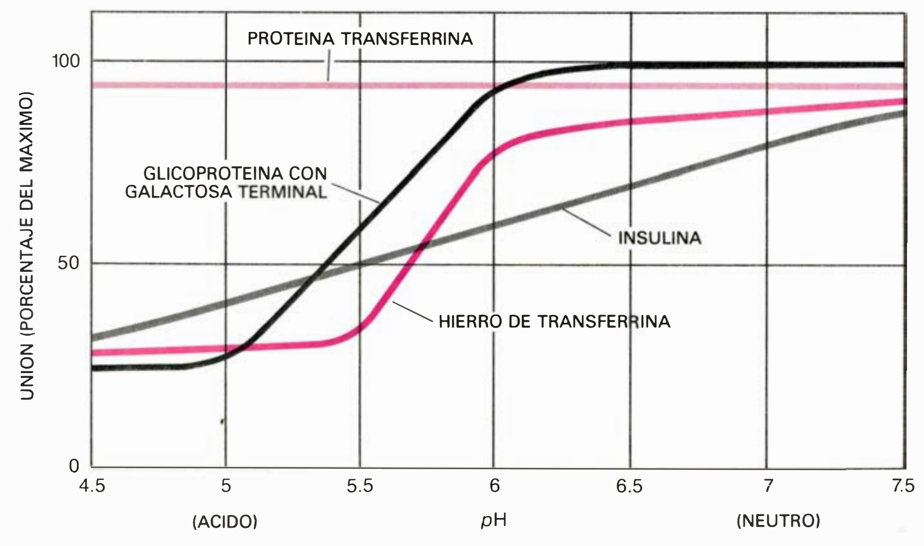
A diferencia de otros ligandos, la ferrotransferrina ni se degrada ni se almacena después de su internamiento, sino que descarga los iones de hierro, que quedan en el interior de la célula, y se secreta rápidamente de la célula en forma de apotransferrina libre de hierro. ¿Por qué se excreta la transferrina de la célula y por qué se desprende de su hierro? En colaboración con Aaron Ciechanover hemos estudiado la endocitosis de la transferrina en cultivos de células tumorales de hígado humano. Las respuestas que hemos hallado sugieren una intervención del pH en la unión de la ferrotransferrina y la apotransferrina al receptor de transferrina. Richard D. Klausner y sus colaboradores, del Instituto Nacional del Cáncer, de los Estados Unidos, alcanzaron conclusiones similares trabajando con cultivos de precursores de glóbulos rojos.

En nuestro laboratorio, del Instituto de Tecnología de Massachusetts, hemos observado que la ferrotransferrina se une ávidamente a su receptor en presencia de un pH neutro del entorno extracelular. Sufre endocitosis y se transfiere a una vesícula prelisosómica ácida, de la misma forma que lo hacen otros ligandos. Gracias a observaciones realizadas hace algunos años quedó establecido que los iones de hierro se disociaban de su transportador transferrina a pH ácido (5,5 o inferior), y hemos observado por nuestra parte que, en efecto, ello ocurre aun cuando la transferrina esté unida a su receptor. Sorprendentemente, la apotransferrina libre de hierro permanece unida a su receptor incluso a pH 5, en contraste llamativo con ligandos tales como LDL, insulina y glicoproteínas con galactosa terminal. De hecho, la afinidad de la apotransferrina libre de hierro a su receptor a pH 5 es la misma que la afinidad de la ferrotransferrina cargada de hierro a pH 7, neutro. Ahora bien, la apotransferrina no tiene afinidad cuantificable por su receptor a pH neutro; si se transfiere súbitamente un complejo apotransferrina-receptor de un entorno acídico a otro neutro, el ligando unido se disocia en segundos.

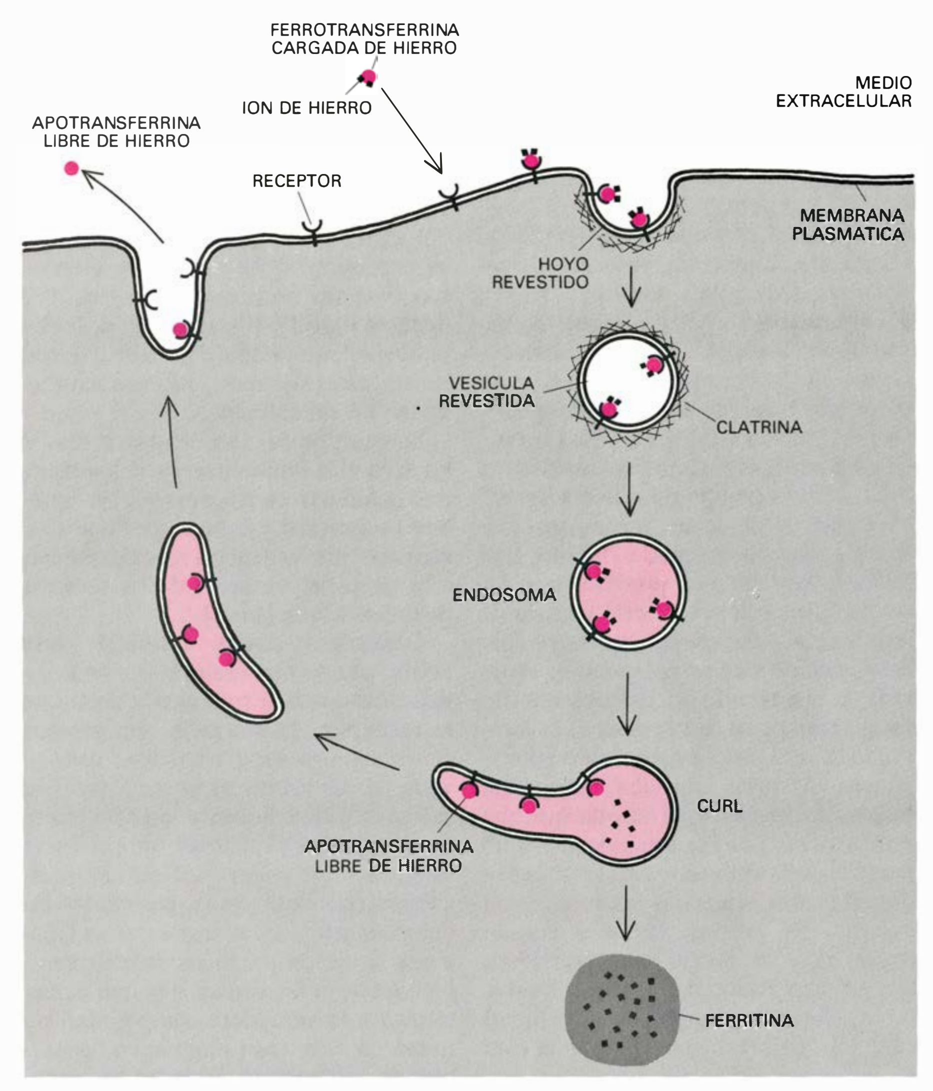
Puede ya trazarse el curso de la transferrina en la endocitosis. La ferrotransferrina, con sus dos iones de hierro, se une al receptor de la superficie celular y el complejo se adentra y se transporta hasta un endosoma; y desde allí hasta una vesícula CURL. Cuando el complejo llega a una vesícula ácida, el hierro se libera. Se transporta (no se sabe cómo) hasta la ferritina, una proteína de almacenamiento de hierro del fluido citoplasmático. El receptor de transferrina vuelve a la superficie de la célula, como ocurre con los demás receptores tras liberar su ligando. Sin embargo, el receptor de transferrina no está libre de su ligando; transporta con él la apotransferrina libre de hierro al exterior de la célula, donde el pH neutro del medio extracelular provoca la disociación de la apotransferrina del receptor. La apotransferrina libre de hierro entra en el torrente sanguíneo, lista para cargar más hierro; el receptor se dispone en la membrana plasmática preparado para engarzar otra molécula de transferrina, por la cual, como se mencionó anteriormente, guarda una alta afinidad en el entorno extracelular neutro.

Todo ello ocurre muy deprisa. Hemos comprobado que cada una de las células de tumor de hígado con las que trabajamos tiene unos 150.000 sitios de unión para transferrina en su superficie o en alguna etapa del reciclaje dentro de la célula. Una célula toma transferrina durante varias horas a una velocidad de 19.000 átomos de hierro, esto es, 9500 moléculas de transferrina por minuto. Dividiendo 150.000 por 9500 se obtiene el tiempo de reciclaje: unos 16 minutos desde la unión de ferrotransferrina a la secreción de apotransferrina. Hemos medido el tiempo requerido para cada paso del ciclo. Por término medio se invierten cuatro minutos en unir en la superficie un ligando de ferrotransferrina a un receptor de transferrina. El complejo ferrotransferrina-receptor tarda cinco minutos en su penetración. Se necesitan otros siete para que se disocie el hierro de la transferrina y para que el complejo apotransferrina-receptor vuelva a la superficie; tras lo cual la apotransferrina se libera del receptor en sólo 16 segundos. El tiempo total transcurrido es de alrededor de 16 minutos.

El proceso global de endocitosis contempla una sucesión de reorganizaciones y fusiones rápidas de las membranas biológicas. Ciertas proteínas de membrana se atrapan selectivamente en un hoyo revestido, mientras que otras sufren exclusión específica. Cada vez que se interna un complejo ligando-receptor, una porción de la membrana plasmática "brota" de la superficie para formar una vesícula revestida. Las vesículas recién generadas se funden, produciendo endosomas y vesículas CURL. Se disocia la membrana de las vesículas CURL; aparentemente, algunas se funden con los túbulos que devuelven los receptores a la superficie y otras se funden con un lisosoma, aportando ligandos para la degradación. A su vez, de los lisosomas pueden "bro-



8. EFECTO DIVERSO DEL MEDIO ACIDO sobre la unión de distintos ligandos. Mientras que la mayoría de éstos se disocian de sus receptores cuando el pH desciende por debajo de 6, la unión de la transferrina a su receptor permanece inalterada. Si se mide independientemente el porcentaje de unión del hierro transportado por ferrotransferrina y el de la propia transferrina, se aprecia que el hierro se disocia del complejo transferrina-receptor a pH inferior a 6, pero la transferrina sigue unida al receptor.



9. DIAGRAMA DEL CICLO DE TRANSFERRINA. Después de la endocitosis del complejo transferrina-receptor, el hierro se libera de la transferrina en el medio ácido de la vesícula CURL, y se transfiere a la ferritina, proteína de almacenamiento de hierro. La apotransferrina libre de hierro permanece unida a su receptor y torna con él a la superficie celular. Cuando el complejo receptor-apotransferrina encuentra el *p*H neutro del medio extracelular, se libera la apotransferrina sin hierro.

tar" partes de su propia membrana en forma de pequeñas vesículas que regresan a la superficie celular y expulsan los productos de digestión.

Las células pueden interiorizar gran parte de su membrana plasmática (hasta el 50 por ciento de ella en una hora), no sólo en endocitosis mediada por receptor, sino también por pinocitosis o fagocitosis. Sería un proceso derrochador para la célula tener que sintetizar continuamente los componentes de la membrana. Es más probable que las porciones de membrana, o sus componentes, tornen a la superficie o al orgánulo al que pertenecen (el lisosoma, por ejemplo). La composición fosfolipídica de la membrana plasmática y de otras membranas es algo distinta, y todas las membranas biológicas tienen dotaciones muy diferentes de proteínas integrales de membrana, de las cuales los receptores de los que hemos tratado constituyen sólo una clase. La membrana plasmática, y todas las membranas biológicas, deben mantener de algún modo su individualidad frente a los procesos continuos de invaginación, fusión y reciclaje.

Debe haber mecanismos que procedan a una continua clasificación de los componentes de la membrana, pero nada sabemos aún de ellos. ¿Cómo se alcanza esa distribución desigual de las proteínas de membrana que logra, por ejemplo, que la membrana plasmática disponga de las permeasas que debe tener y en el lisosoma aparezcan las enzimas apropiadas para el bombeo de protones? ¿Cómo se orientan las vesículas de un orgánulo a otro? ¿Hay señales específicas de clasificación, o moléculas parecidas a receptores, en la superficie externa de cada vesícula? Para comprender con detalle la endocitosis mediada por receptor habrá que responder, primero, a cuestiones de ese tenor.

Ciencia y sociedad

De la complejidad en la ciencia

Bajo el lema "La ciencia y la práctica de la complejidad" se desarrolló en Montpellier, del 9 al 11 de mayo de este año, un simposio organizado por la Universidad de las Naciones Unidas, cuya oficina central radica en Tokio. Intervinieron directamente veintitantas personas de diferentes especialidades, aunadas por un mismo interés en el tema enunciado. La diversidad de enfoques puso de manifiesto la preocupación común por un complejo de cuestiones, más o menos filosóficas, que aparecen en distintas disciplinas. Sin embargo, esta misma aproximación pluridisciplinar impidió el examen profundo de algunas cuestiones fundamentales y las obvias preconcepciones filosóficas dificultaron a veces la comunicación.

Los temas o títulos de las distintas sesiones dan idea de cómo se había concebido la organización interna del simposio, cuyo contenido se puede deducir también de los nombres de los presentadores de las distintas ponencias que configuraron cada sesión: *1*) Comprensión de la complejidad (Danzin, Klir, Le Moigne); *2*) La complejidad expresada por relaciones dinámicas (Costa de Beauregard, Hägerstrand, Pribram); *3*) El aprender y el significado (Boulding, Luhmann, Atlan); *4*) Ecosistemas (Margalef, Holling, Aida); *5*) Aplicaciones (Allen, Voge); *6*) Estructuras sociales (Dupuy, Zeleny, Giarini) y *7*) Aproximaciones a la complejidad (Laborit, Morin, Prigogine). Se presentaron también algunas breves comunicaciones complementarias y hubo amplio tiempo para la discusión.

Bastantes personas reconocen una complejidad en el mundo externo, que se suele generar por procesos que incluyen bifurcaciones y cierta medida de indeterminación; parece que la posible existencia de esta complejidad es aceptada de buen grado por físicos y naturalistas. Para otros, sin embargo, la complejidad está sólo en el observador y guarda bastante relación, con esta forma de concebirla, el aceptar que una buena medida de la misma puede ser el número de símbolos utilizados para describir una situación, generalmente externa. No falta quien ve el mundo como algo irreductible, o considera que es vana ilusión de la ciencia buscar un fundamento simple a la realidad compleja (Morin); punto de vista éste que puede ser elocuente, pero muy opuesto a cualquier fe en la coherencia del mundo físico y en la posibilidad de su expresión a través de una ciencia. Las distintas posiciones entroncan con debates filosóficos que duran milenios y, obviamente, nadie esperaba que con las retóricas del momento se desvanecieran las dificultades.

La cuestión fundamental se acota en un área más limitada, si ordenamos en dos columnas cierto número de nombres, conceptos y expresiones que aparecieron, por lo general repetidamente, a lo largo del conjunto de las sesiones [*véase cuadro adjunto*].

Cualquiera puede continuar esta doble lista de manera más o menos satisfactoria, con lo cual queda claro que se reconoce, de una parte, un proceso generativo de cada situación, que no viene dada arbitrariamente como un capricho de los dioses, y, de otra parte, lo que podemos llamar un producto final, en cuya generación se ha inmiscuido cierta dosis de lo que parece indeterminación, de manera que se llega a una situación prácticamente irrepetible. Aunque las situaciones que caracterizan a la verdadera complejidad no faltan en el mundo inorgánico (pensemos en la superficie de cualquier planeta), son mucho más irreductibles dentro de los dominios de la vida, de sus producciones y de sus epifenómenos. Wittgenstein pudo decir que, en lógica y en matemática, el proceso equivale al resultado; pero en los dominios donde campea lo que llamamos complejidad hay sustancialmente un relleno histórico que impide que la descripción del proceso, limitada, encierre la infinidad de pequeñas variantes que son posibles, introducidas como ruido (¿creador?) y conducentes a lo que me complace llamar lo barroco en la naturaleza.

Desde el punto de vista de empresas concretas como, por ejemplo, el estudio de los ecosistemas planctónicos, se puede visualizar bien el paso de los procesos hidrográficos y biológicos hacia el estado final representado por una distribución en el espacio en un momento dado. A partir de este motivo de distribución deberíamos o desearíamos poder reconstruir los procesos generadores. Probablemente es inútil empeñarse en utilizar toda la información existente en la distribución final observada, ya que sólo podemos aspirar a

CONCEPTOS RELACIONADOS CON LA COMPLEJIDAD	
PROCESO	ASPECTO FINAL DE UN DESARROLLO HISTORICO
DETERMINISMO	INDETERMINISMO IMPREDECIBILIDAD INESTABILIDAD
ORDEN NEWTON, LAPLACE	DESORDEN
FINITUD	INFINITUD
LEY FISICA QUE ELIMINA EL TIEMPO	TEMPORALIDAD (EVENTUALMENTE CREADORA)
COMPLICACION DE ARTEFACTOS	COMPLEJIDAD DE LA NATURALEZA
SIMETRIA	ASIMETRIA
SIMPLICIDAD ORGANIZADA	COMPLEJIDAD DESORGANIZADA
"LEY" CIENTIFICA	DESCRIPCION DIFICIL DE ABREVIAR, COMO LOS CONTORNOS DE UN MAPA
INTELIGENCIA	REDUNDANCIA, HISTORIA, BARROCO
ESTILO LITERARIO	LO QUE SE DICE
ARGUMENTO, O COMO ACABA LA NOVELA	ESTRUCTURA DE LA REDUNDANCIA EN UNA NOVELA
GENOTIPO	FENOTIPO
PLANCTON	BENTOS O COMUNIDADES TERRESTRES

Lista de expresiones más frecuentes del simposio celebrado en Montpellier sobre "La ciencia y la práctica de la complejidad".

utilizar el "estilo" de la distribución en la reconstrucción. Dentro del estado actual de la ciencia oceanográfica, el detalle de la distribución no dice gran cosa, quizá por la razón que el lirio atigrado le dio a Alicia: "Hablamos cuando hay alguien con quien merece la pena hablar".

Aún me gusta más otro ejemplo que, creo, profundiza en los temas inherentes a la teoría y práctica de la complejidad: la visión. Nuestro ojo capta más información de la que puede conducir nuestro nervio óptico; una fracción relativamente mayor de la misma procede de áreas donde paramos atención, y el resto se rellena, con más incertidumbre, por la mente. Si hay que fiarse de los sueños, esta interpolación puede ser de mucho detalle, pero en estado de vigilia está continuamente filtrada por su contraste con información directa, por limitada que ésta sea. Este ejemplo presenta, en toda su viveza, los problemas de la complejidad y de la información, como organización y como conocimiento, e incluso como especulación y ensueño.

Otro tema de fondo implícito, y a veces explícito en varios de los debates, es la distinción entre un soporte operador y una memoria, obvia en los artefactos, que se desvanece, no obstante, cuando se pasa a los organismos, de los que puede decirse que son todo memoria. El paso conceptual de una a otra situación se puede explorar aplicando con cierta dosis de imaginación los autómatas o máquinas de Turing. Para el biólogo, los sistemas más interesantes son los formados por elementos reproducibles, de distintas clases (individuos de distintas especies, células de diversos tipos), cuya combinación constituye un sistema de orden superior, dentro del cual aquellos subsistemas cobran significado. Este problema no es específico de la biología: las palabras son elementos reproducibles que forman el lenguaje, dentro del que cobran sentido y que viene a ser el canal, donde "fluye" la información o comunicación. Si el ecosistema se toma como referencia, las especies y los individuos tienen relación con él (adaptación, etcétera); si el individuo se toma como referencia, toda la actividad mental se puede ver como un canal cerrado, o abierto sólo hacia el infinito, pero desconectado de los otros organismos que forman la sociedad o el ecosistema, e independiente de sus procesos de generación y evolución. Tal sobrevaloración del individuo, presente, por ejemplo, en los escritos de Maturana y Varela, y que contrasta con un punto de vista más clásico

o biológico, no estuvo ausente en el simposio. Zeleny pasó de comentar sistemas capaces de autoorganización en los dominios de la física y de la química, mencionando incluso las "plantas artificiales" de Leduc y su éxito efímero o reprimido, a tratar de los sistemas humanos más complejos, con muchas citas de precursores diversos. Cuando le pregunté por qué dejaba de lado a Darwin y no consideraba intermediarios más apropiados sistemas genuinamente biológicos, pues el tema, a pesar de las exageraciones del darwinismo social y de la biosociología, no podía ser tabú, ni siquiera para un humanista, me contestó que a él le interesaba sólo el individuo, en los términos de autorreferencia de Varela. Disquisiciones semejantes llevaron más de una vez a comparar las actitudes filosóficas del Este y del Oeste, expresando cierta simpatía hacia la actitud oriental, aunque también se hizo notar que la civilización occidental se ha mostrado como la única crítica consigo misma y capaz de renovarse.

Pribram expuso su opinión según la cual ciertas actividades mentales y la memoria compartirían propiedades de los hologramas, especulando que la experiencia mística puede basarse en la proyección de un holograma cerebral total. En su opinión la teoría de la comunicación de Shannon no tiene que ver con la información, sino con señales, que, como voces de mando (se originó durante la guerra), piden obediencia. Todo el mundo está de acuerdo en que hay una explosión de documentos, pero no una explosión de información. Atlan, en una comunicación muy precisa, propone como algoritmo de la complicación el número de instrucciones que hay que dar a una máquina de Turing en relación con artefactos. Hace notar que la observación no puede identificar el paso de información de un nivel a otro, que sería esencial para la autoorganización. Presentó un sistema o modelo reticulado de interacciones, semejante al "juego de la vida", en el que se generan motivos que gozan de cierta estabilidad en su estilo. Sin embargo, este modelo, como otros semejantes, no tiene en cuenta las relaciones ternarias y su inestabilidad, que resultan esenciales en los sistemas biológicos y ecológicos.

Holling ilustró diferentes conceptos con el estudio de las fluctuaciones de poblaciones de un insecto defoliante en bosques norteamericanos, haciendo ver el papel de la interacción entre diversos conjuntos de variables que característicamente cambian a velocidad diferente. Me interesó vivamente su representación de entidades sobre un plano cuyas dimensiones se definen como velocidad de cambio (rápido, lento) y como intensidad de acoplamiento o interacción (débil, intenso), porque resulta muy comparable con una representación que he usado con referencia al fitoplancton. Las especies que siguen la estrategia evolutiva designada por la r (oportunistas) corresponden a un cambio rápido y a una conexión débil; las estrategas de la K, a velocidad lenta y a acoplamientos fuertes. Estas ideas se pueden extender a sistemas económicos. Allen presentó los modelos suyos y de sus colaboradores sobre la evolución de las ciudades; pero más sugerente es su utilización de las pajaritas de papel como ejemplo de que, partiendo de piezas de papel semejantes, por una serie de dobleces sucesivas (bifurcaciones), se consiguen estructuras cualitativamente distintas y discontinuas. Este modelo es aplicable al desarrollo embrionario y a la evolución.

Para Voge, el modelo de Simon según el cual si hay una jerarquía en el ensamblaje, y la selección es simultánea en los distintos niveles, la evolución se acelera, es aplicable a sistemas económicos. Pero esta estrategia de construcción es menos pertinente en relación con la naturaleza, donde no hay problemas de finalidad ni prisas en el desarrollo de estructuras. Los economistas ven como novedad el riesgo de una vulnerabilidad no asegurable (Giarini) y, aunque se menciona el problema de la rigidez de la oferta y su poder e inflexibilidad como resultado de la organización, quizá no se aprecia la falta de una realimentación (*feedback*) apropiada, ausencia que aumenta exageradamente el poder (y la información) de la oferta en relación con la demanda. Prigogine ofreció una presentación clásica de la generación histórica de sistemas complejos, haciendo ver la intervención de amplificadores en la generación de ciertas regularidades, cuyo reconocimiento no permite, sin embargo, la predicción. A la pregunta de la "oposición" (Zeleny, Morin) de si creía que detrás de la complejidad se hallaba alguna simplicidad que *merezca la pena* investigar, Prigogine contestó que sí; la diferencia entre un sí y un no en este caso define la divergencia más profunda que puedo ver en la aproximación al problema de la complejidad. Cualquier decisión se ve con menos pasión por el naturalista, o por el que practica cualquier ciencia en general, acostumbrado a utilizar instrumentos imperfectos para aproximarse a la comprensión de una realidad exterior, y que sabe que para conocerla no puede fiar totalmente en un órgano seleccionado como factor de supervivencia. A mayor abundamiento, es imposible que una parte de un sistema entienda el sistema más amplio del que forma parte. Toda reseña es parcial y cada uno lo cuenta desde su punto de vista. El simposio tuvo, naturalmente, muchos otros aspectos interesantes sobre los que soy incapaz de opinar.

Boulding, para quien la ciencia es una mutación evolucionista de la cristiandad, además de insistir en la importancia de las diferencias individuales, y criticar la matemática por conducir a un punto de vista reduccionista, se entretuvo poniendo en verso la crónica del simposio. La riqueza de la civilización, verdadero paradigma de la complejidad, estuvo muy presente en la reunión. Los participantes franceses, con su propensión a dejar caer adjetivos muy bien colocados, se dejaron llevar a veces por la inercia retórica, contrastando con la mayor sequedad y aparente falta de imaginación de los anglosajones, sazonado todo por la chispa vienesa (von Förster asistió también al simposio). No sé si el simposio habrá hecho avanzar mucho la ciencia de la complejidad, pero resultó aleccionador para todos y en muchos aspectos.

El tema de la complejidad y de los sistemas que vemos como capaces de conseguirla por sí mismos goza actualmente de popularidad y hay otras reuniones internacionales anunciadas para fechas próximas sobre el mismo asunto. Es posible que el simposio reseñado tenga ciertas consecuencias o continuidad de una u otra forma. Más cerca de nosotros, en Barcelona, hay anunciadas unas jornadas sobre sistemas autoorganizativos, para enero de 1985, de cuya organización cuida la cátedra de termodinámica y fisicoquímica de la Escuela Técnica Superior de Ingenieros Industriales de Barcelona. (Ramón Margalef.)

Miniacerías en España

Si bien es verdad que el hierro es el metal que con mayor profusión se ha venido utilizando desde la edad del Hierro, unos 1500 años a. C., también es cierto que el procedimiento de obtención de su aleación más empleada, el acero (aleación de hierro y de carbono), ha experimentado, a lo largo del tiempo, distintas y sustanciales modificaciones. A modo de síntesis, se puede afirmar que los métodos de obtención de hierro y de acero, a través de la his-

toria, han sido los siguientes: procedimientos directos, hornos altos y, actualmente, las miniacerías.

Los procedimientos directos de obtención de hierro y de acero consistían, en esencia, en la reducción en pequeños hornos de la mena concentrada de hierro con monóxido de carbono obtenido "in situ" por combustión del carbón. Después de unas cuantas horas de operación se lograba una masa porosa e impura más o menos acerada. Se purificaba y se compactaba por forja, o deformación en caliente. Así se conseguían los perfiles que consumía la industria derivada: cerrajería, utillajes, clavos, armas, etcétera. Estos procedimientos desaparecieron cuando se descubrió un carbón más resistente a elevada temperatura que los existentes en aquel entonces (hacia 1750). Con la aparición del carbón de coque, fue posible utilizar hornos alargados y de gran capacidad, pues la resistencia mecánica del carbón de coque encendido era suficiente para aguantar la carga del horno alto, formada por la mena, el carbón y el fundente.

El horno alto, nacido pues con el descubrimiento del carbón de coque, produce un producto intermedio entre la mena y el acero. Se trata del arrabio o aleación base hierro con elevados porcentajes de carbono, manganeso, silicio, fósforo y azufre. El arrabio se descarbura y se purifica en los hornos de las acerías integrales, donde se convierte en acero.

Actualmente la tecnología del horno alto está en declive debido a varios factores, entre los cuales cabe citar: aumento del parque de chatarra en los países más desarrollados, empobrecimiento gradual de ciertos tipos de menas, encarecimiento del coque metalúrgico y necesidad de instalar los centros productores de acero próximos a los centros de consumo. Estos factores han inducido la aparición de las miniacerías: industrias siderúrgicas que utilizan fundamentalmente chatarra para producir acero en hornos eléctricos de arco o de inducción. Es decir, mientras la siderurgias integrales emplean el arrabio del horno alto, las miniacerías consumen chatarra. De esta manera, países carentes de menas de hierro, tales como el Japón y Suiza, se han convertido en importantes potencias siderúrgicas.

Contemplando este panorama tecnológico surge la pregunta: ¿cómo está la siderurgia española? Actualmente España produce mensualmente más de un millón de toneladas métricas de acero, lo que le sitúa en el lugar número trece en la lista de países productores de acero. Este acero se fabrica en las siderurgias integrales existentes en el norte (Asturias y Vizcaya), además de la de Sagunto, y en las miniacerías que, a partir de la mitad del siglo XX, se han desarrollado con mucho éxito en Cataluña y en Euskadi. En cuanto a tonelaje, la producción siderúrgica de las miniacerías es comparable con la de las siderurgias integrales. En cuanto al tipo de acero, las integrales suelen fabricar acero al carbono y las miniacerías producen aceros semialeados y aleados. Un producto típico de las integrales es el perfil en forma de *T* o de doble *T* obtenido por colada continua en el oxiconvertidor, mientras que un producto típico de la miniacería lo constituye el redondo corrugado; ambos perfiles se emplean en la construcción.

La aparición de las miniacerías españolas ha venido motivada fundamentalmente por la necesidad de la importación de menas de hierro. No obstante, España necesita también importar chatarra, sobre todo de Francia, Reino Unido y Estados Unidos de América. Si bien este aspecto de la balanza de pagos está bastante equilibrado debido a la exportación de perfiles de acero. La siderurgia española está al día. (Pedro Molera.)

Miniacerías

Mientras grandes siderurgias que parten de mineral de hierro tienen dificultades, las pequeñas que se abastecen de chatarra y comercializan localmente una limitada gama de productos gozan de esplendor

Jack R. Miller

Desde 1977, las 10 empresas norteamericanas que integran el grupo de las llamadas "Grandes Acerías" han cerrado, parado, transferido o vendido 20 plantas o secciones de planta, disminuyendo su capacidad de producción en unos 10 millones de toneladas métricas, que corresponden al 8 por ciento de los 123 millones de toneladas métricas de producción anual. Medidas que han acarreado también la reducción de plantillas de la fuerza laboral ocupada en la industria del acero, que ha pasado de 453.000 en 1979 a 247.000 a finales de 1982, es decir, una disminución del 45 por ciento en tres años. A lo largo de 1982, la industria del acero perdió, en su conjunto, 3200 millones de dólares, mucho más que en cualquier ejercicio precedente. El objetivo de las Grandes Acerías es la reestructuración, descrita por David M. Roderick, presidente de US Steel Corporation, como un "estado de autoliquidación acelerado". La reestructuración incluye cierre de talleres, postergación o cancelación de proyectos, consolidación de la producción y goteo de los productos.

Contrasta con lo anterior el auge experimentado por el sector industrial de las llamadas miniacerías, desde 1960. Ese año había 10 o 12 miniacerías que cubrían aproximadamente el 2 por ciento del mercado estadounidense del acero. A comienzos del año en curso, medio centenar de ellas abastecían del 15 al 18 por ciento del mercado. Su capacidad de producción total (estimada en plantas que pueden producir hasta 800.000 toneladas métricas anuales) se aproxima a los 14 millones de toneladas métricas anuales. Los empresarios tienen puesta su mirada en la expansión de plantas y mercados, así como en la ampliación del abanico de productos fabricados.

Aunque las Grandes Acerías atribuyen muchos de sus problemas a un aluvión de importaciones de acero a bajo coste, algunos observadores creen que esas empresas no están exentas de culpa por su falta de decisión a la hora de adoptar nuevas tecnologías y cierta complacencia en las técnicas de mercado. En una reciente entrevista concedida al *Financial Times*, el presidente de una miniacería argumentaba: "Las grandes compañías mantuvieron precios altos y desestimaron mercados cuando las ventas estaban en alza. Ahora, las compañías persisten equipadas con plantas y talleres anticuados que ya lo eran desde hace años. No les queda otra salida que la reconversión".

Por pertinente que sea el consejo, las Grandes Acerías quizá lo rechacen al venir de quien viene. Por otro lado, aunque las miniacerías hayan adquirido una importancia creciente y se hayan afianzado en el sector del acero desde 1970, siguen operando en ámbitos marginales. Fabrican varilla estructural, barra y conformados ligeros con bajo valor añadido de manufactura que son destinados a la venta en el mercado local de la construcción, compitiendo así con el acero importado. Su materia prima es principalmente chatarra de acero local que se funde en hornos eléctricos. Por tanto, si exceptuamos los costos de adaptación de nueva técnica a sus procesos, su presupuesto representa varias veces menos que un modelo entero de acería nacional.

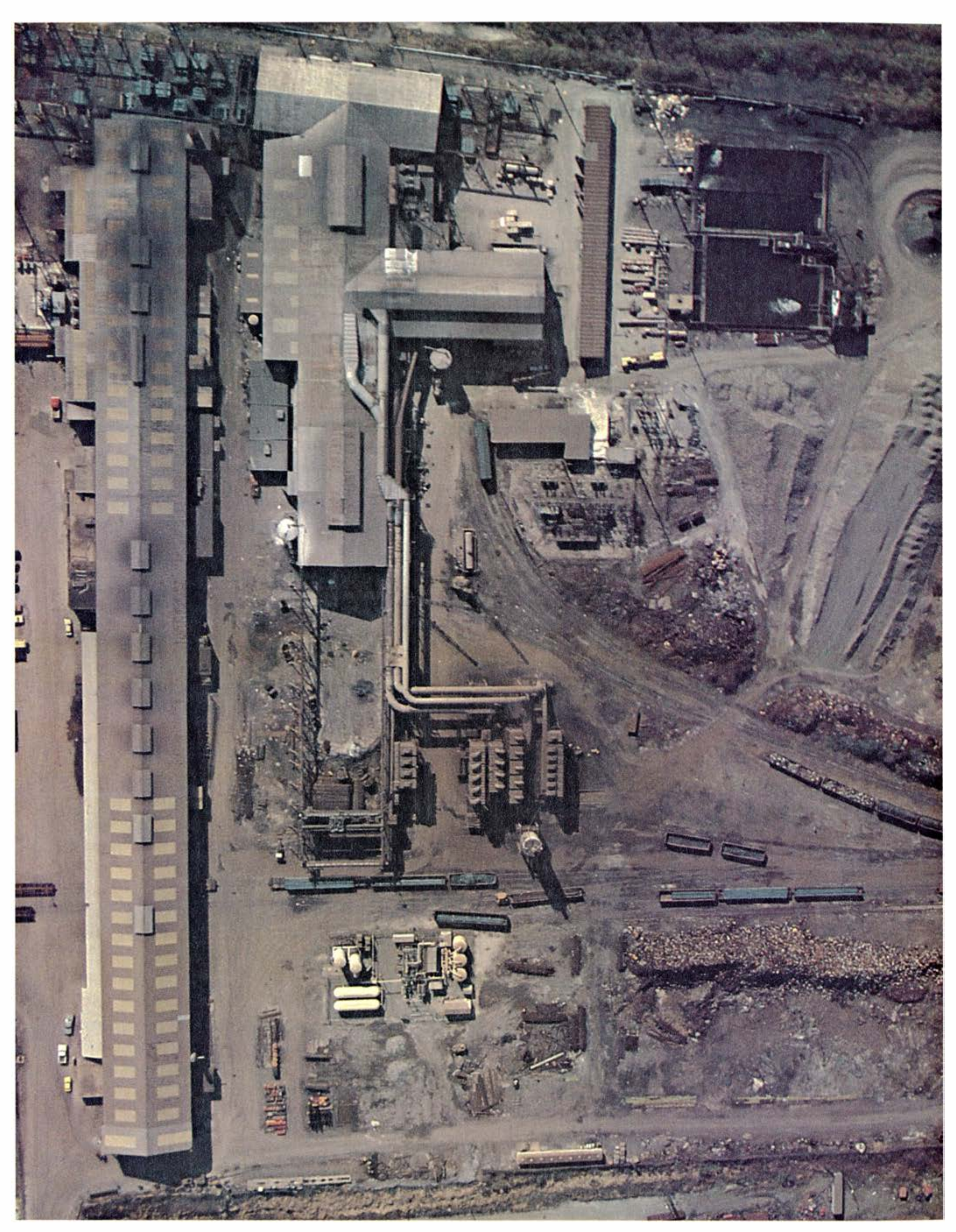
Una acería capaz de iniciar la fabricación con mineral de hierro y finalizarlo produciendo múltiples tipos y tamaños de conformados requiere una mayor inversión por tonelada de capacidad productiva que una miniacería. Los datos sobre costes de capital fueron recogidos en 1980 por la Oficina de Tasación Tecnológica del Congreso de los Estados Unidos. El rango estimado es de 154 a 320 dólares de 1980 por tonelada métrica de capacidad anual para miniacerías y de 956 a 1500 para plantas integrales (llamadas así porque en ellas se obtienen varios conformados de acero a partir de mineral de hierro).

El primer paso de la mayoría de las operaciones integradas es la extracción del mineral; el segundo, la preparación de las materias primas. El mineral de baja ley, que antes de la Segunda Guerra Mundial solía despreciarse, ahora se beneficia. Su contenido de hierro se aumenta hasta valores entre el 65 y 72 por ciento mediante la separación del material de desecho, como arena, impurezas metálicas y arcilla. Los minerales concentrados se mezclan, para endurecerlos posteriormente en forma de pequeñas bolas (pelets) de 1,0 a 1,5 centímetros de diámetro. Los finos del mineral se sinterizan, es decir, se funden hasta obtener una masa porosa que después se trocea en piezas de 2,5 a 5,0 centímetros. Es frecuente la instalación de plantas de peletización cerca de la mina; los pelets se transportan en tren o barco hasta la acería, procedimiento que viene facilitado por el mejor empaquetamiento de los pelets frente a los sinterizados. Las plantas de sinterizados constituyen, sin embargo, una sección típica de la acería integral.

En las plantas integrales, la preparación de materias primas abarca también la trituración de la caliza y la obtención de coque por la combustión parcial de carbón en ausencia de oxígeno. Tanto el coque como la caliza se trituran hasta alcanzar el tamaño deseado.

El tercer paso corresponde a la obtención de hierro. Los minerales preparados, la caliza y el coque se mezclan y cargan en el horno alto por su parte superior, mientras que, por la base, se sopla una corriente de aire caliente. Este aire quema el coque generando monóxido de carbono que sube por el interior del horno y alcanza el mineral que desciende lentamente. El gas caliente funde y reduce el mineral. Como resultado se obtiene arrabio líquido.

En el cuarto paso, el arrabio es transportado y afinado en un horno de



1. MINIACERIA de Tampa. Pertenece a Florida Steel Corporation. La materia prima de la planta es chatarra de acero comprada en el lugar. La chatarra se funde en hornos eléctricos, que están en el segundo edificio de la izquierda, y se cuela en forma de palanquillas en una máquina de colada continua en el mismo edificio. Las palanquillas se laminan hasta obtener varillas de refuerzo para hormigón en el tren de laminación instalado en el largo edificio de la izquierda. El agua para la refrigeración de la colada continua proviene del depósito rectangular situado en la parte central superior. El colector de humos, que separa el polvo, está situado en el edificio perpendicular al que contiene el horno eléctrico. El material azul, visible en algunos de los vagones de ferrocarril, son varillas acabadas y palanquillas que se embarcan para su laminación en otra parte. La planta produce unas 254.000 toneladas.

afino, que suele ser un horno básico con soplado de oxígeno o un horno eléctrico. Durante esta etapa, se eliminan los elementos desechables y se ajustan las cantidades finales de los elementos deseados, particularmente carbono, que se mantiene entre 0,3 y 1,7 por ciento. El acero fundido se vierte en una cuchara y de ésta es colado en lingoteras de fundición, donde solidifica en forma de lingote o de planchón grueso ("slab") que pesan alrededor de 60 y 100 toneladas, respectivamente.

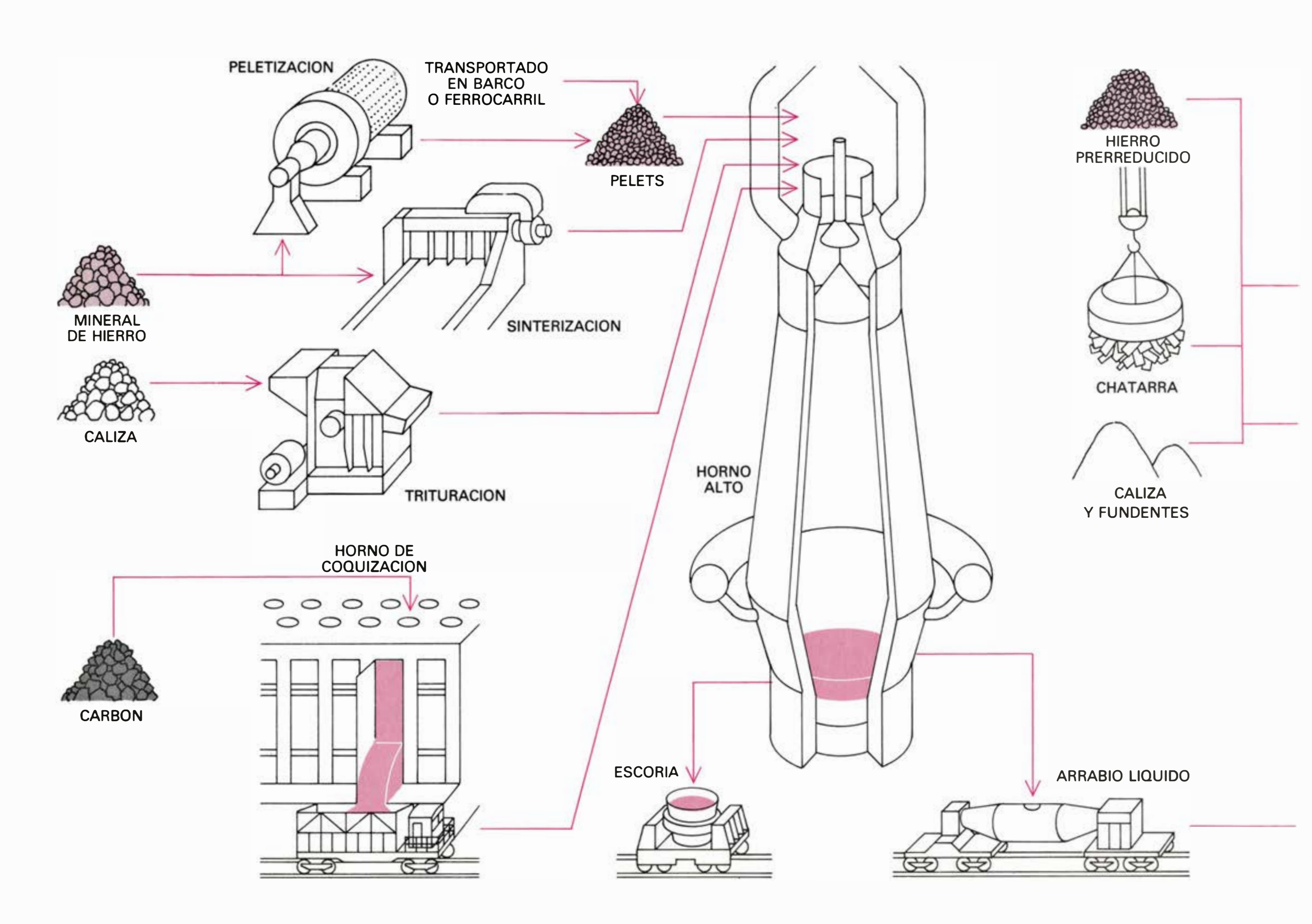
Los lingotes y los planchones se desmoldean tan pronto como es posible y se envían a una nave exterior para su almacenaje o tratamiento superficial o bien al taller de laminación. Allí los lingotes se introducen en un horno de fosa y los planchones ("slabs") en un horno de precalentamiento, donde ambos alcanzan una temperatura uniforme. Se cargan en el tren laminador de desbaste, donde reciben el conformado preliminar antes de pasar a los trenes de laminación intermedia y final.

Durante los años setenta, algunas acerías integrales instalaron máquinas de colada continua, que reciben el acero fundido directamente de la cuchara produciendo palanquillas, lingotes gruesos ("blooms") o planchones ("slabs") destinados al tren de laminación; se eliminan así las lingoteras y las etapas de deslingotado y precalentamiento. Desde hace 15 años, la colada continua de palanquilla es el procedimiento normal en las miniacerías.

En los trenes de laminación los planchones se laminan hasta obtener: plancha para buques, grandes depósitos industriales o tuberías soldadas de gran diámetro; plancha fina para la fabricación del automóvil; banda estrecha para pequeños productos manufacturados; hojalata y plancha galvanizada. Los lingotes cuadrados ("blooms") y palanquillas grandes son laminados para obtener perfiles estructurales de gran sección. De las palanquillas pequeñas se obtienen perfiles estructurales ligeros, barras para refuerzo del hormigón, varillas, alambre y tubo no soldado.

Las miniacerías reciben tal nombre no tanto por su tamaño cuanto porque sus procesos sólo abarcan parte de una acería integral. Las acerías integrales se levantan para producir de tres a diez millones de toneladas anuales de lingotes y alcanzar así la economía de escala que permita reducir los costes de producción. En Estados Unidos, la mayoría de las miniacerías tienen una capacidad de unas 250.000 toneladas métricas anuales y sólo algunas alcanzan 600.000 toneladas o más. En Hawai hay una planta de 50.000 toneladas; otros países poseen instalaciones de la mitad de capacidad.

Joseph C. Wyman, analista de este sector industrial perteneciente a Shearson/American Express, Inc., se ha ocupado de la economía de las miniacerías. En 1980 escribía que las miniacerías "no debían invertir en talleres para tratar la materia prima ni en transporte",



2. ACERIA TRADICIONAL, denominada también planta integral porque inicia el proceso con mineral de hierro y lo termina con productos de acero acabados o semiacabados. En el horno alto se cargan mezclas del mineral en terrones y pelets (y quizá sinterizados) para obtener arrabio. Con la operación de afino se elimina el carbono del baño metálico para fabricar acero, que tiene un contenido de carbono de 1,7 por ciento o menor. El arrabio, chatarra, caliza y fundentes se cargan en el convertidor (los dos tipos aquí descritos son un horno de arco eléctrico y un horno básico al oxígeno) donde el carbo-

ni habían de obtener el arrabio como producto intermedio. "Para conseguir su materia prima confían en las inversiones de otros: chatarreros y compañías de reciclaje. El conjunto de productos que fabrican también es 'mini'; sólo producen un limitado rango de perfiles y tamaños con un estrecho abanico de productos."

Sin necesidad de gran inversión de capital, propia de las plantas integrales, las miniacerías gozan de una mayor libertad de reinversión en nueva tecnología. Así es como han conseguido conquistar mercados que las Grandes Acerías habían dejado en manos de la importación.

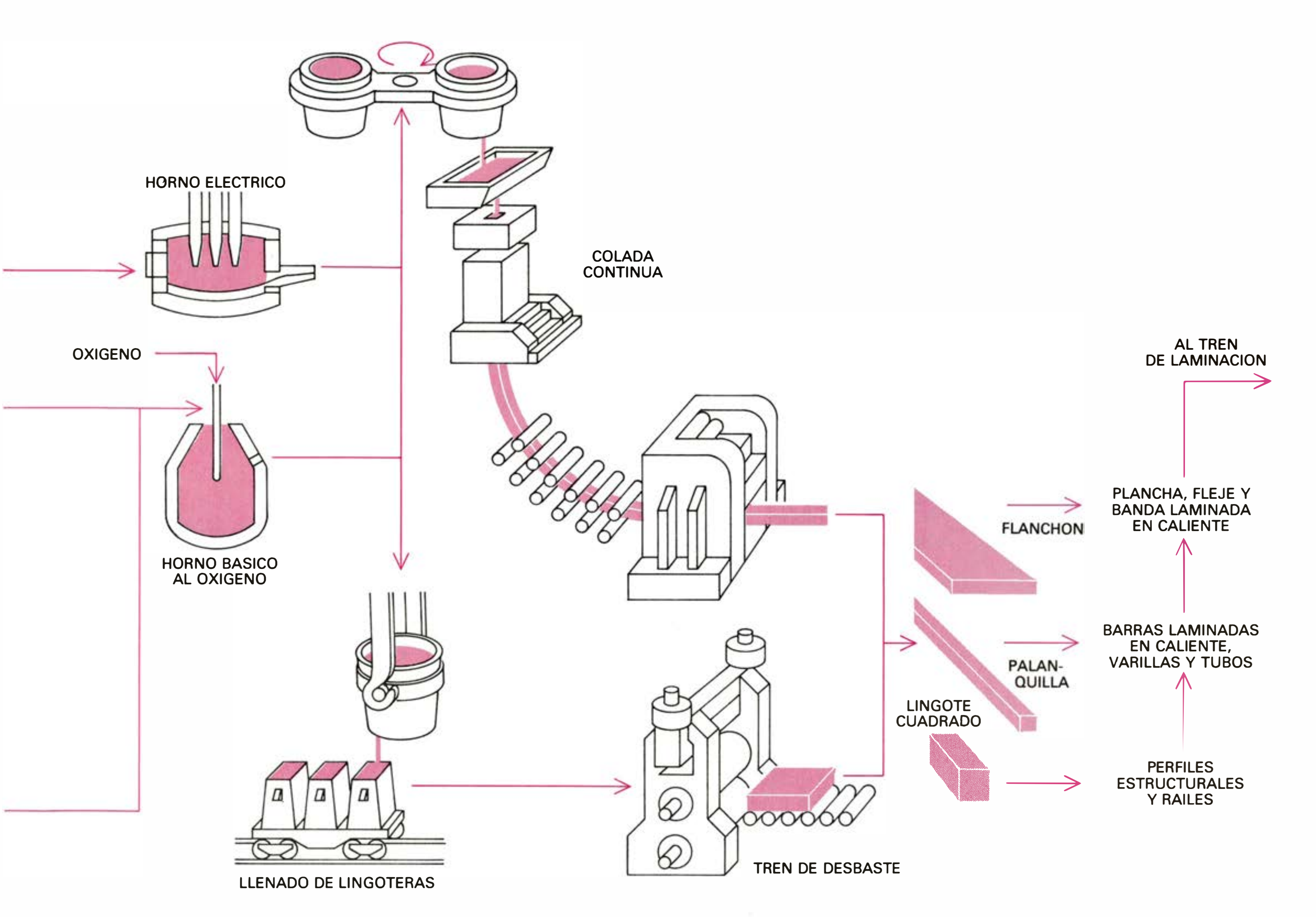
En una miniacería, el primer paso es fundir la chatarra de acero en un horno de arco eléctrico. El acero fundido entra en la máquina de colada continua, de donde sale en forma de lingotes cuadrados ("bloom") o palanquilla continua, que se corta con la longitud conveniente y se carga en el tren de laminación para conformarla en perfiles; fabrican, pues, una gama de productos laminados limitada.

El horno, la máquina de colada continua y el tren de laminación forman parte de cualquier planta integral grande, pero aquí se describe su utilización en una miniacería. La parte principal del horno eléctrico moderno es la vasija, de tres a ocho metros de diámetro, recubierta interiormente con material refractario. A través del techo del horno entran tres electrodos de grafito. Estos electrodos están conectados a un transformador que permite aplicar de 130 a 700 kilovolt-ampère por tonelada métrica de carga. Según una norma propuesta en 1982 por el Instituto Internacional del Hierro y del Acero, se clasifican como hornos de ultrapotencia las unidades que alcanzan los 700 kilovolt-ampère. El calor generado en el sistema se debe a los arcos eléctricos producidos entre los electrodos y la chatarra de acero que contiene la vasija.

El conjunto formado por el techo y los electrodos puede separarse de la vasija. El horno se carga con la chatarra contenida en una gran cubeta suspendida por una grúa puente instalada en el taller. A la carga suele añadirse arrabio, o más recientemente hierro prerreducido, a fin de diluir el efecto de las impurezas, tales como cobre, plomo, cinc y cromo presentes en la chatarra.

Puede emplearse oxígeno para acelerar la fusión de la carga y para eliminar las impurezas de carbono que haya en el baño metálico. Cuando tal ocurra, se inyecta en el interior del horno mediante una lanza refrigerada con agua. El horno eléctrico trata de 80 a 100 toneladas en cada etapa (período entre dos coladas) con una duración de 100 a 180 minutos, que corresponde a una velocidad de 45 a 65 toneladas métricas por hora. El consumo de energía eléctrica oscila alrededor de 550 kilowatt-hora por tonelada métrica.

Al final del proceso se retiran el techo y los electrodos; la vasija bascula



no, manganeso y silicio se reducen hasta los niveles especificados mediante oxidación controlada y donde se eliminan los contaminantes fósforo y azufre. El acero se cuela en lingoteras para obtener planchones ("slabs"), palanquillas o lingotes cuadrados ("blooms") que posteriormente son laminados para obtener los productos finales. Muchas acerías integrales instalan máquinas de colada continua de planchones para reemplazar el procedimiento de colada en lingotera. Las plantas integrales contratan un mayor número de empleados de varias categorías y fabrican más acero que una miniacería típica.

para vaciar la carga de acero líquido en una cuchara sostenida por la grua puente. En este momento se inicia la colada.

En las primeras miniacerías, se realizaba la colada en pequeñas lingoteras en forma de palanquilla. Una vez solidificado el acero, las palanquillas, deslingotadas, se enviaban directamente al horno de precalentamiento previo al laminado. En 1962 se introdujo una modificación importante: la instalación de máquinas de colada continua que reemplazó a la colada en lingoteras.

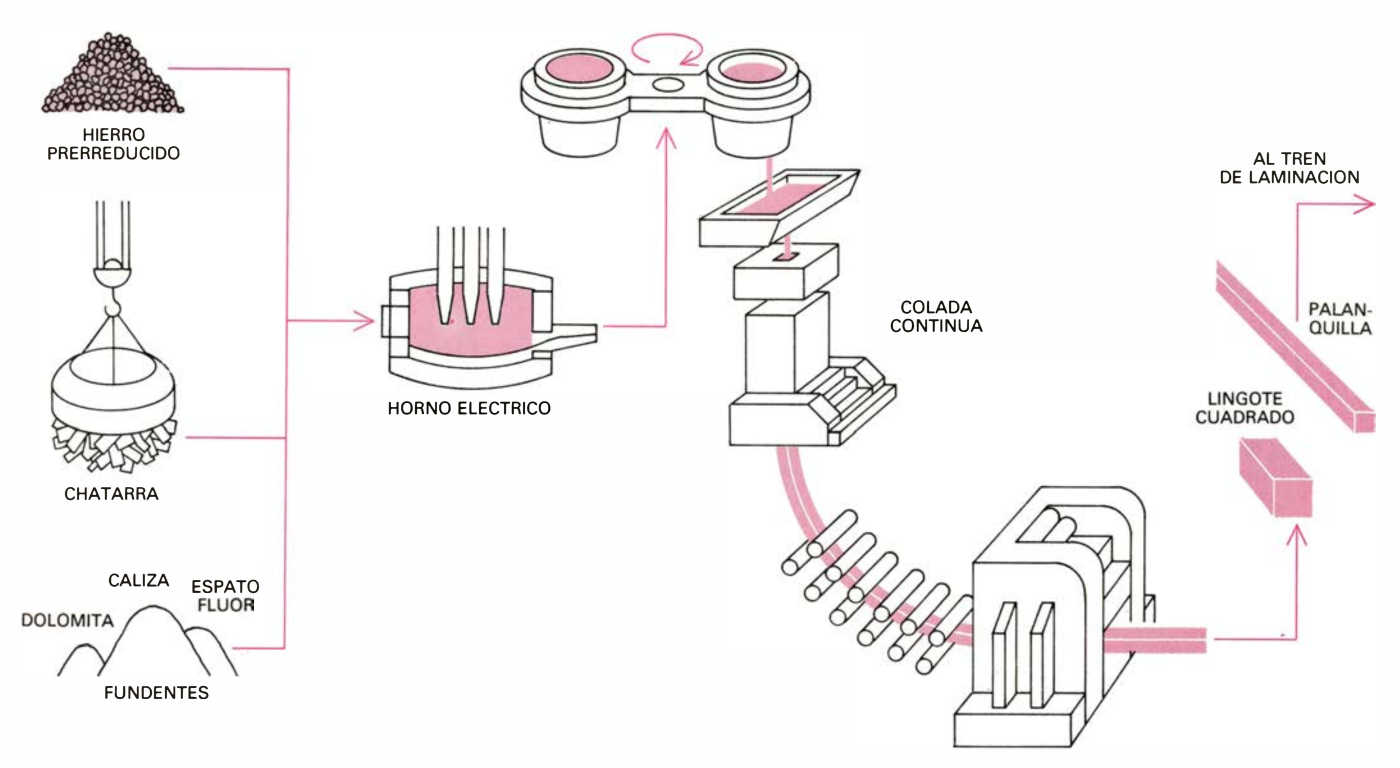
En este tipo de máquinas, el acero llega de la cuchara a un distribuidor elevado que actúa de embudo. Este distribuidor tiene una abertura en la parte inferior, a cuyo través el acero líquido cae en un molde de cobre refrigerado con agua. El metal y el molde descienden unos 25 milímetros actuando el molde de soporte mientras se forma una fina capa de acero solidificado. Entonces el molde asciende bruscamente, separándose de la capa sólida ya formada y vuelve a su posición inicial para dar soporte a una nueva cantidad de acero líquido.

La palanquilla, aún pastosa, prosigue su descenso a través de un conjunto de guías soporte refrigeradas con agua pulverizada que la curvan hasta alcanzar un transportador a ras del suelo. La palanquilla continua, ya totalmente sólida, atraviesa un enderezador que actúa con rodillos niveladores. En algunas plantas, la palanquilla continua se corta a longitud conveniente. Un avance reciente consiste en alimentar en continuo el horno de precalentamiento previo al laminado y, más reciente aún, alimentar en derivación el horno.

Los productos fabricados en la sección de laminación de una miniacería son sencillos y tienen una gama limitada. Normalmente incluyen alambrón, barra corrugada para hormigón y una variada forma de barras de sección redonda, cuadrada o rectangular. También se fabrican vigas ligeras de sección en I o en T, ángulos (con sección en ángulo recto) o canalones (con sección en *U*). Si la parte más larga de la sección no llega a 75 milímetros se habla de sección ligera y de sección pesada cuando supera esa cifra. Las barras fabricadas con acero al carbono laminadas en caliente reciben el nombre comercial de barra, mientras que la denominación acero aleado se aplica en aquellos casos donde están presentes pequeñas cantidades de manganeso, cromo, níquel u otros elementos. Tales productos, de fabricación creciente en las miniacerías, se inscriben en la calidad SBQ (barras de calidad especial).

La calidad de barras y varillas ha mejorado constantemente en los últimos 10 años, debido, sobre todo, a su creciente papel en las miniacerías. Los controles dimensionales son más precisos, la calidad superficial de los productos ha mejorado y el control de la temperatura del acero y la presión de los rodillos de laminación se regula por ordenador. Por otra parte, la velocidad de laminación ha aumentado considerablemente alcanzándose en la actualidad velocidades de 100 metros por segundo durante la etapa final de la laminación de varillas de 5,5 milímetros de diámetro, ritmo que dobla, de lejos, el que se lograba diez años atrás.

Un atento repaso al funcionamiento de dos miniacerías nos revelará que, en general, se han mostrado abiertas a las nuevas técnicas, se ha maximizado el rendimiento del proceso, se ha estudiado el alojamiento de la planta y las técnicas de mercado necesarias. Nos detendremos primero en la planta que Raritan River Steel Co. instaló en Perth Amboy, Nueva Jersey, que entró en funcionamiento en marzo de 1980. El lugar dista unos 32 kilómetros del centro de Nueva York y unos 120 kilómetros de la ciudad de Filadelfia. Permite que un camión a una velocidad de 90 kilómetros por hora enlace en menos de cuatro horas con los clientes de Boston, Worcester, Providence, Al-



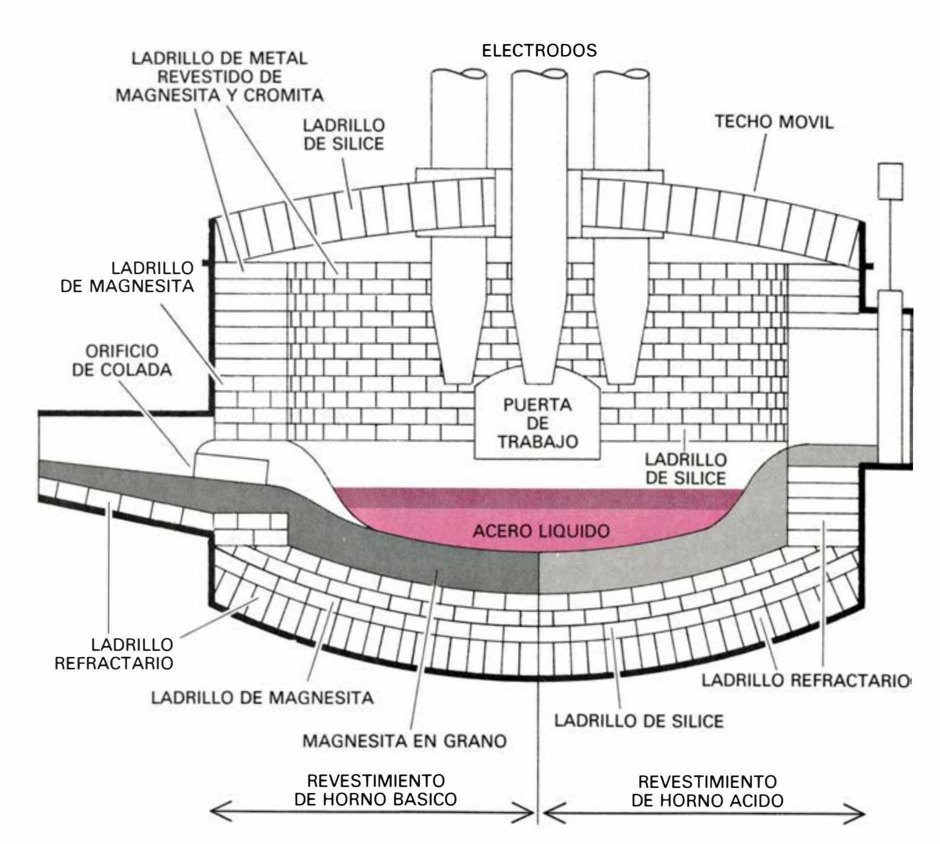
3. ASI FUNCIONA UNA MINIACERIA. Se ahorra la etapa de obtención de hierro de las plantas tradicionales y se empieza con chatarra de hierro, fundentes y, en algún caso, hierro prerreducido. La chatarra se funde en el horno eléctrico, se vacía en las cucharas y se transporta hasta la máquina de colada continua. De la máquina de colar el acero emerge en forma de palanquilla o lingote cuadrado continuo, que se corta a longitud conveniente y se envía al tren de laminación, donde normalmente se produce un grupo de acabados restringido. Estos incluyen varillas, barras y conformados estructurales para construcción ligera. En los últimos años las miniacerías han aumentado su gama de productos: tubo, plancha, fleje fino y plancha fina.

bany, Binghamton, Scranton, Harrisburg y gran parte del área de Baltimore-Washington. El conjunto de esta región, habitada por 55 millones de personas, constituye cerca del 25 por ciento de la población de los Estados Unidos y cerca del 25 por ciento del consumo total de acero. Esta planta solamente fabrica varilla.

Los talleres que componen la fábrica son un ejemplo de la nueva tecnología aplicada en las miniacerías. Su único horno eléctrico de seis metros de diámetro produce 135 toneladas métricas de acero a partir de chatarra. El acero líquido se carga en una máquina de colada continua de cinco canales, que fabrica palanquillas de 130 milímetros de lado por 16 metros de longitud. La agitación electromagnética del caldo metálico permite obtener acero al carbono de gran calidad y microestructura uniforme. Una innovación de esta máquina de colada permite sustituir la sección de molde de un canal sin detener la producción en los restantes. El tren de laminación compuesto por 25 grupos de rodillos aumenta la longitud y reduce la sección de las palanquillas hasta obtener varillas de 5,5 a 12,7 milímetros de diámetro.

La capacidad nominal de esta planta es de 550.000 toneladas métricas de varilla por año. La planta, construida durante 1978 y 1979, costó 110 millones de dólares. Emplea trabajadores no sindicados en una zona donde las organizaciones sindicales han gozado tradicionalmente de una gran fuerza. El aprendizaje de los operarios ha merecido especial atención. Al iniciarse la producción, se estableció como productividad óptima 1,36 hombre-hora por tonelada métrica; durante 1983 la productividad alcanzó un valor de 1,63 hombre-hora por tonelada en condiciones rutinarias.

El segundo ejemplo corresponde a las cinco miniacerías de Florida Steel Corporation. Una de ellas, que fue la primera planta construida en Florida, se instaló en Tampa en 1958 para producir 3000 toneladas métricas mensuales de barra corrugada para hormigón, con el propósito de satisfacer las necesidades de la propia compañía. Las barras gozaron pronto de una amplia aceptación en el mercado local; la compañía se convirtió en fabricante. Durante 25 años la empresa prosperó, amplió la planta de Tampa y edificó otras cuatro similares: en Charlotte, Carolina del Norte, en 1961; en Indiantown, Florida en 1970; cerca de Jacksonville en 1975 y en Jackson, Tennessee, en



4. EL HORNO ELECTRICO constituye la parte central de la miniacería. Alcanza hasta 8 metros de diámetro y contiene 150 toneladas métricas de metal fundido. Tres electrodos de grafito penetran a través de otros tantos agujeros del techo. Cuando la potencia está conectada, se genera gran cantidad de calor. Saltan arcos eléctricos de un electrodo a las piezas de chatarra de acero, que constituyen la carga metálica del horno, y de la carga al electrodo siguiente. La carga se funde, formando un baño metálico cubierto por una capa de escoria. El revestimiento refractario difiere para procesos ácidos o básicos. En el proceso básico habitual las reacciones químicas (que involucran la caliza de la escoria, la magnesita de los refractarios y el metal fundido) eliminan carbono, silicio, manganeso, fósforo y azufre del baño y producen sales fundidas. Se retira el conjunto formado por el techo y los electrodos; el horno bascula y vacía el acero.

1981. Los costes de la planta de Indiantown excedieron a los previstos y fue clausurada en 1982. La capacidad de laminación de las plantas de Tampa y Charlotte es de 250.000 toneladas métricas anuales, de 320.000 la de Jacksonville y de 350.000 la de Jackson.

La compañía se basó en distintas razones para arriesgarse en la expansión, que ilustran tres supuestos fundamentales del método de operación de una miniacería. Las plantas de Tampa y Charlotte se construyeron para aprovechar las posibilidades que ofrecía el mercado local y la existencia de mano de obra y chatarra: esa es la idea subyacente a toda miniacería. Las de Indiantown y Jacksonville pretendían cubrir una demanda creciente y establecer mercados cerca de las áreas de competidores potenciales: éste es el concepto de vecindad. Pero sigue siendo una miniacería que busca establecerse en una zona donde se espera un crecimiento de la demanda de acero. La planta de Jackson se construyó para satisfacer la demanda existente y cubrir un esperado crecimiento del consumo de barras y corrugados. Esta miniacería se limita a fabricar unos productos específicos y hacer frente a un determinado mercado: éste es el concepto de mercado.

La planta de Jackson incorpora un tipo avanzado de colada continua llamado colada secuencial o colada continua en continuo. El horno eléctrico y el taller de colada continua se diseñaron de suerte que el tiempo necesario para obtener suficiente metal fundido para llenar una cuchara fuera menor que el necesario para pasar el acero a través de la máquina de colar. Se vacía así una serie de cucharas en los canales de colada de forma continua, sin que deba pararse la producción cuando una cuchara esté limpia.

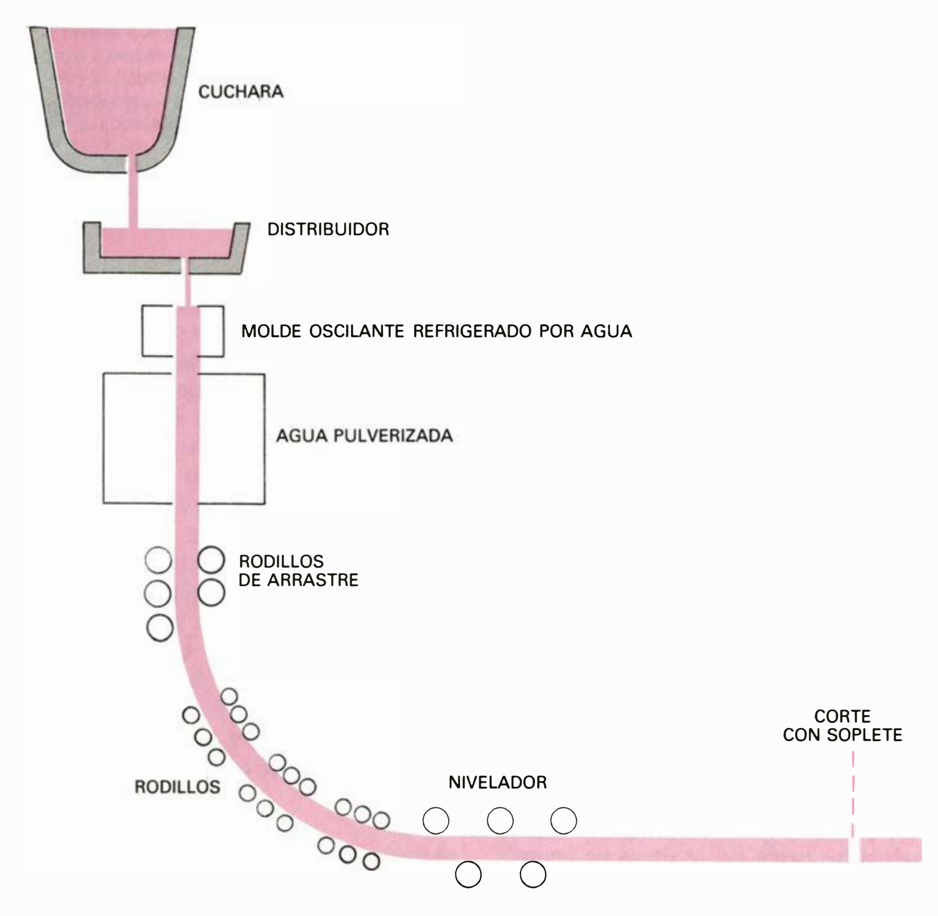
Puesto que las miniacerías dependen del suministro de chatarra de acero, cabe preguntarse si este suministro podrá mantenerse, conocido el creciente papel que las miniacerías tienen en la industria del acero. James W. Brown y Richart L. Reddy, de la división de sistemas de electrodos de la Union Carbide Corporation, examinaron, hace un

año, esta cuestión y llegaron a la conclusión de que el suministro de chatarra en los Estados Unidos parecía ser "adecuado para soportar las necesidades domésticas y niveles de exportación futura". Hans Mueller, de la Universidad estatal de Tennessee Central, cree que "la gran disponibilidad de chatarra en todo el país" tiene importancia para la instalación de miniacerías equipadas con hornos eléctricos en los estados del oeste y del sur, que "han sido olvidados por las grandes compañías". Según estadísticas publicadas por el Instituto de la Chatarra de Hierro y Acero, la chatarra alcanzaba los 620 millones de toneladas métricas a pricipios de 1982, en los Estados Unidos. Una reserva de esta magnitud, junto con las adiciones anuales, bastaría para hacer frente a la demanda predecible durante varias décadas.

La chatarra es una materia prima de bajo coste, si la comparamos con el mineral de hierro o el arrabio. Según un estudio de Union Carbide realizado el pasado año, donde se coteja el coste de chatarra por tonelada de acero producido en una miniacería con el coste de metal fundido producido en horno alto en una planta integral, las miniacerías ganan con una diferencia en torno a los 100 dólares.

La miniacería aventaja también a las plantas integrales en lo concerniente a coste de mano de obra y energía. Aquellas emplean normalmente personal no sindicado. La dirección ha propiciado unas buenas relaciones con los trabajadores mediante planes de reparto de beneficios, cursos de preparación, programas de promoción y consultas sobre las decisiones. Aunque los casos varían, puede decirse, como aproximación, que el coste total de empleados por tonelada de acero producido oscila entre 75 y 100 dólares en las miniacerías y de 195 a 295 en las plantas integrales.

Robert W. Bouman, de la Bethlehem Steel Corporation, calcula, en su análisis de las necesidades energéticas, un consumo de 5,6 millones de kilojoule para obtener una tonelada de acero en un horno eléctrico a partir de chatarra en una miniacería y de 16,96 millones de kilojoule en un horno alto durante un proceso básico con oxígeno en una planta integral. "Desde un punto de vista energético concluye, es deseable el uso de chatarra para producir acero."

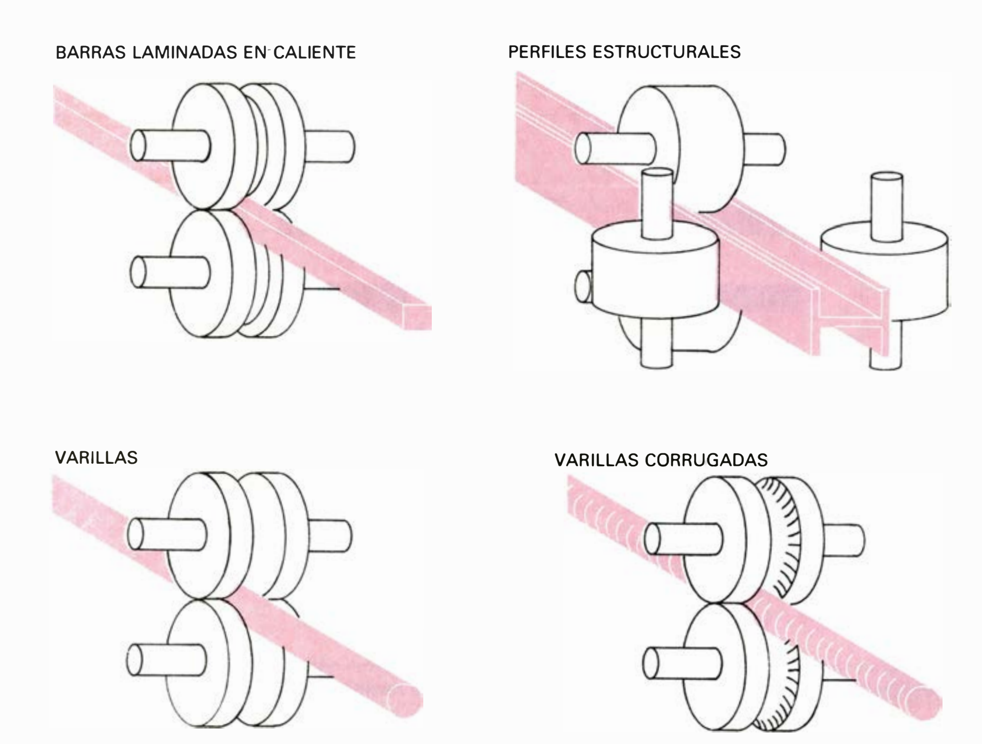


5. COLADA CONTINUA, método normal de colada en las miniacerías. El acero fundido transportado al taller de colada en una cuchara sostenida por una grúa puente se vierte en el distribuidor que, actuando como embudo, alimenta de líquido los moldes de cobre refrigerados por agua. El metal y el molde bajan una pequeña distancia permitiendo que solidifique una fina capa de acero. Entonces, la oscilación separa el molde del acero solidificado y vuelve a su posición inicial. La palanquilla de acero sigue bajando a través de pulverizadores de agua. Cerca del suelo se curva hasta alcanzar la posición horizontal. Se cortan secciones que se envían al horno de precalentamiento previo a la laminación. Una práctica reciente consiste en enviar directamente la palanquilla desde la máquina de colada hasta el tren de laminación.

Las siguientes predicciones personales sobre el curso de la siderurgia en los próximos siete años (1984-1990) están basadas en mi propia relación industrial y en recientes discusiones con los interesados en el sector. Los cambios que se suceden con rapidez en la industria no permiten estimar períodos más largos. Mis predicciones no son exactamente pronósticos calculados, sino juicios que se apoyan en las tendencias observadas.

Espero, en primer lugar, que el consumo aparente de acero en los Estados Unidos aumente de 123 millones de toneladas métricas en 1985 a 127 millones en 1990. (Se llama consumo aparente de acero a la producción neta más el comercio internacional neto.) La previsión está basada en una tasa de crecimiento anual de 0,62 por ciento, que podría resultar modesta en exceso. También estimo que el nivel de importaciones, que en los años recientes ha sido del 19 al 22 por ciento del mercado interior de los Estados Unidos, descenderá hasta aproximadamente el 16,5 por ciento en 1985, para permanecer así hasta 1990. Parte de este descenso en las importaciones resultará del aumento de la actividad de las miniacerías. Habiendo ganado una gran parte de este mercado en su presente línea de productos (90 por ciento de conformado ligero, 73 por ciento de barras corrugadas, 55 por ciento de varillas y 30 por ciento de barras laminadas en caliente en 1981), las miniacerías van extendiendo su línea de productos hasta cubrir todo tipo de producto de acero: tubería y tubo, plancha, hojalata, lámina, banda. Como antes, las miniacerías realizarán su penetración en el mercado a expensas de las importaciones y de las plantas integrales.

El acero producido por el método del horno eléctrico acumuló el 28 por ciento de la producción total estadounidense durante 1980; preveo un aumento hasta el 32 por ciento en 1985 y al 36 por ciento en 1990. La chatarra para el horno eléctrico probablemente se conseguirá en cantidades adecuadas a precios que oscilarán entre 118 dólares por tonelada en 1985 a 126 dólares por tonelada en 1990. Precios promedio que



6. ENTRE LOS PRODUCTOS FABRICADOS tenemos barras, varillas lisas y corrugadas y perfiles estructurales ligeros. Las barras pueden ser planas, redondas o cuadradas (llamadas así por la forma de su sección). En algunas plantas, los productos se laminan hasta obtener barras, que se reducen a alambre.

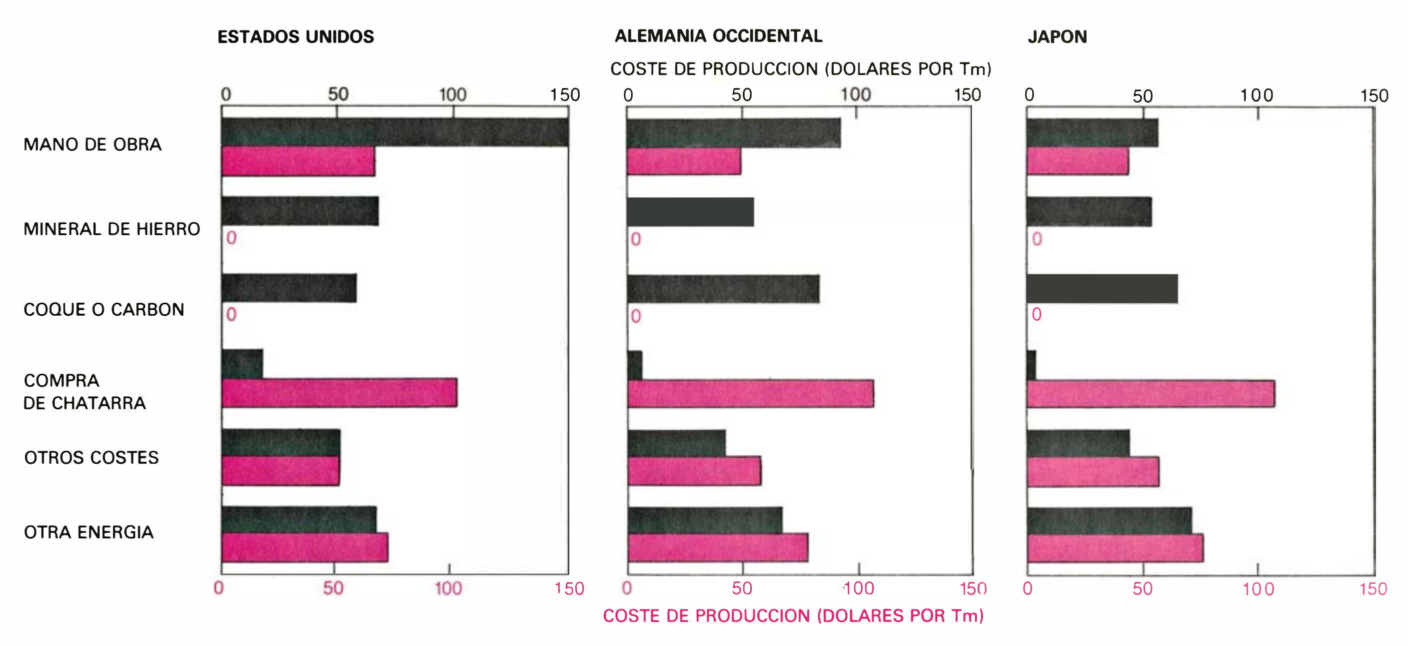
no se sobrepasarán, debido a la restricción competitiva impuesta por la disponibilidad de hierro prerreducido como carga alternativa para hornos eléctricos. En este proceso, el mineral de hierro es reducido a estado metálico sólido por calentamiento sin fusión. El resultado es un hierro altamente reducido que sirve de materia prima para la producción de acero.

Las miniacerías esperan seguir creciendo. Ahora bien, la cada vez mayor competencia que les harán las Grandes Acerías reestructuradas y otras miniplantas modernizadas impedirá que mantengan su tasa de crecimiento por encima de los valores de la última década. La capacidad de las miniacerías, cifrada en más de 14 millones de toneladas métricas en 1983, aumentará constantemente hasta 1990. Estimo que las miniacerías podrán cubrir del 20 al 24 por ciento del mercado de los Estados Unidos, es decir, rendir una producción total de unos 22 millones de toneladas métricas en 1985 y 26,5 millones en 1990. También parece verosímil que muchas amplíen su gama de productos y la calidad de aceros que fabrican.

Hay que esperar que se registren avances tecnológicos, principalmente en miniacerías. Nuevas técnicas aplicadas en el taller de fusión permitirán obtener entre 60 y 80 toneladas de material fundido en menos de una hora. Recientes innovaciones, ya probadas en algunos talleres de fusión equipados con hornos eléctricos, se introducirán en la mayor parte de los procesos de fabricación. Entre estas "prácticas secundarias" señalaré el afino en cuchara, mediante el cual la chatarra se funde rápidamente en un horno eléctrico con la aplicación de la máxima potencia del mismo; el metal fundido se vacía rápidamente en la cuchara donde se afina. El procedimiento permite reservar el horno eléctrico para su mejor aplicación: la fusión a alta potencia.

Otra innovación a considerar será la del procedimiento de la escoria esponjosa que se desarrolló en México hace ya 25 años. En este proceso se añade carbón y coque al horno al principio de la fusión; las reacciones resultantes producen la formación de una capa (de 30 a 120 centímetros de espesor) de escoria borboteante de baja densidad en la parte superior del baño. Esta escoria esponjosa envuelve y protege los electrodos, facilita la transferencia de calor desde los electrodos al metal, reduce las salpicaduras y restringe al mínimo el deterioro del material refractario.

De las innovaciones más eficaces del último decenio destacaré la utilización de paneles refrigerados por agua en la construcción de las paredes y techo de los hornos eléctricos. Esta práctica per-



7. COSTES DE PRODUCCION por tonelada métrica de alambrón producido en miniacerías comparados con plantas integrales en 1981. En negro se representan los costes para una planta integral y, en color, los correspondientes a una miniacería típica. Las miniacerías no tienen gastos de carbón.

mite eliminar un 75 por ciento de la pared y techos refractarios tradicionales y, por tanto, ahorrar en coste de producción.

Estas y otras técnicas están encaminadas a intensificar el rendimiento, disminuir los costes de producción y mejorar la calidad. Muchas de ellas ya se utilizan en plantas de construcción reciente. Es de presumir que las nuevas técnicas entren de pleno en la mayoría de las miniacerías en los próximos cinco años.

En la colada continua también se prevén importantes innovaciones. Existe interés en conseguir máquinas horizontales de colada continua para salvar el riesgo de derrames durante el traslado de grandes cucharas llenas de acero fundido suspendidas de las grúas puente a una altura de más de un metro del nivel del suelo. También evitaría la necesidad de grandes y costosas estructuras que alojan las máquinas de colada continua vertical. Pero dudo que las máquinas horizontales sustituyan a las verticales en un futuro próximo.

Con unos leves retoques, la velocidad de laminación de muchos de los trenes existentes puede incrementarse en más del 50 por ciento. Más importante aún, algunas miniacerías podrán fabricar en 1990 planchones ("slabs") de 25 a 40 milímetros de grosor y de 600 a 1000 milímetros de anchura mediante colada continua vertical y laminarlos directamente (es decir, pasarlos directamente a las etapas de acabado del tren de laminación).

Dos grandes miniacerías de los Estados Unidos, por lo menos, practican ya de una manera regular la laminación directa. De la colada continua salen dos líneas de palanquilla (algunas veces sólo una) y se transfieren simultáneamente en continuo al tren de laminación, donde son laminadas sin el precalentamiento usual. (La temperatura de las palanquillas se mantiene uniforme durante su recorrido mediante calentadores de inducción instalados entre los primeros grupos de rodillos.) En 1985, esta práctica se aplicará probablemente en media docena o más de miniacerías.

Los últimos años de la década de 1990 verán el principal avance técnico: la colada directa de fleje de 3 a 5 milímetros de espesor, que constituye la materia prima para muchas industrias de transformación que producen gran variedad de productos acabados. En ese próximo decenio quizá se asista también a la introducción de antorchas de plasma y hornos de fusión de plasma en algunas miniacerías de los Estados Unidos. El plasma es un gas ionizado que consiste en un número aproximadamente igual de iones y electrones. Se forma al pasar combustible gaseoso comprimido a través de un arco eléctrico generado entre dos electrodos refrigerados con agua. El combustible se fragmenta e ioniza con gran desprendimiento de calor. Las técnicas de plasma se utilizan ya en varios países. En Suecia y Japón están en servicio antorchas de plasma para calentamiento auxiliar; en Alemania del Este dos hornos de plasma llevan produciendo acero aleado de alta calidad desde hace tres años.

El afino en cuchara, la colada en continuo, la laminación directa de palanquillas y la colada y laminación directa de planchones delgados y flejes establecen el marco para la producción en continuo de una línea completa de productos de acero en operaciones ininterrumpidas, donde la materia prima entre por uno de los extremos y los productos acabados salgan por el otro. Se ha demostrado la posibilidad de acoplamiento entre el horno eléctrico y la colada continua, así como entre la unidad de colada y el tren de laminación. Es de esperar que antes de acabar el siglo se pueda fabricar acero en continuo en una o más miniacerías norteamericanas.

Finalmente, aunque la participación de las miniacerías en la producción de acero en los Estados Unidos aumente desde el 20 por ciento en 1985 al 24 por ciento en 1990 y no ceda en esa línea ascendente, no creo que el sector reemplace a las plantas integrales. Es verosímil que la producción de acero en miniacerías alcance un máximo entre el 32 y el 40 por ciento, probablemente cerca del 34 por ciento, antes de 1995. La razón de poner ese límite reside en que los productos de acero, incluida la chatarra que sustenta la producción de las miniacerías, proviene de mineral de hierro. Sin el sector de las plantas integrales, cuyos hornos altos inician la cadena de mineral a acero, a productos de acero, luego, y chatarra por último, las miniacerías se verían obligadas a cerrar antes de diez años.

Parásitos que cambian el comportamiento de su patrón

Al hacerlo, logran que el patrón sea más vulnerable a la depredación por parte de su siguiente hospedador. Entre ellos se cuentan ciertos acantocéfalos, que infestan a cochinillas, devoradas luego por pájaros

Janice Moore

Uno de los habituales recursos literarios de la fantasía científica es el de pintar parásitos extraterrestres que invaden un hospedador humano y le obligan a realizar sus designios, al tiempo que se multiplican y se extienden a otros desventurados terrícolas. Pero la idea de que un parásito pueda alterar el comportamiento de otro organismo no es mera fantasía. Ni tan siquiera se trata de un fenómeno raro. Basta con acercarse a un lago, un campo o un bosque y observar atentamente.

Durante largo tiempo, la mayoría de los ecólogos que se interesan por el componente animal de las comunidades vivas han preferido concentrarse en los animales de vida libre, y dejar el estudio de los parásitos a los parasitólogos. Se sabía que muchos parásitos tienen complicados ciclos biológicos: pasan los primeros estadios de su vida en un animal (el patrón u hospedador intermediario) y alcanzan la madurez en otro (el patrón definitivo). Gradualmente se fueron acumulando indicios de que algunos parásitos no esperan en actitud pasiva que la casualidad les permita alcanzar su destino final, sino que poseen métodos para aumentar la probabilidad de que el primer patrón caiga presa del segundo. Pueden conseguirlo cambiando simplemente el tamaño o el color del patrón intermediario, potenciando sus rasgos. Por contra, en otros casos alteran el comportamiento del animal, aumentando su vulnerabilidad. En este sentido, ya en 1931 Eloise B. Cram, del Departamento de Agricultura de los Estados Unidos, observó que saltamontes infestados con las larvas del nemátodo *Tetrameres americana* eran presa fácil de los pollos, en los que este gusano vive de adulto; las larvas se enquistan en los músculos del saltamontes y menguan su vivacidad.

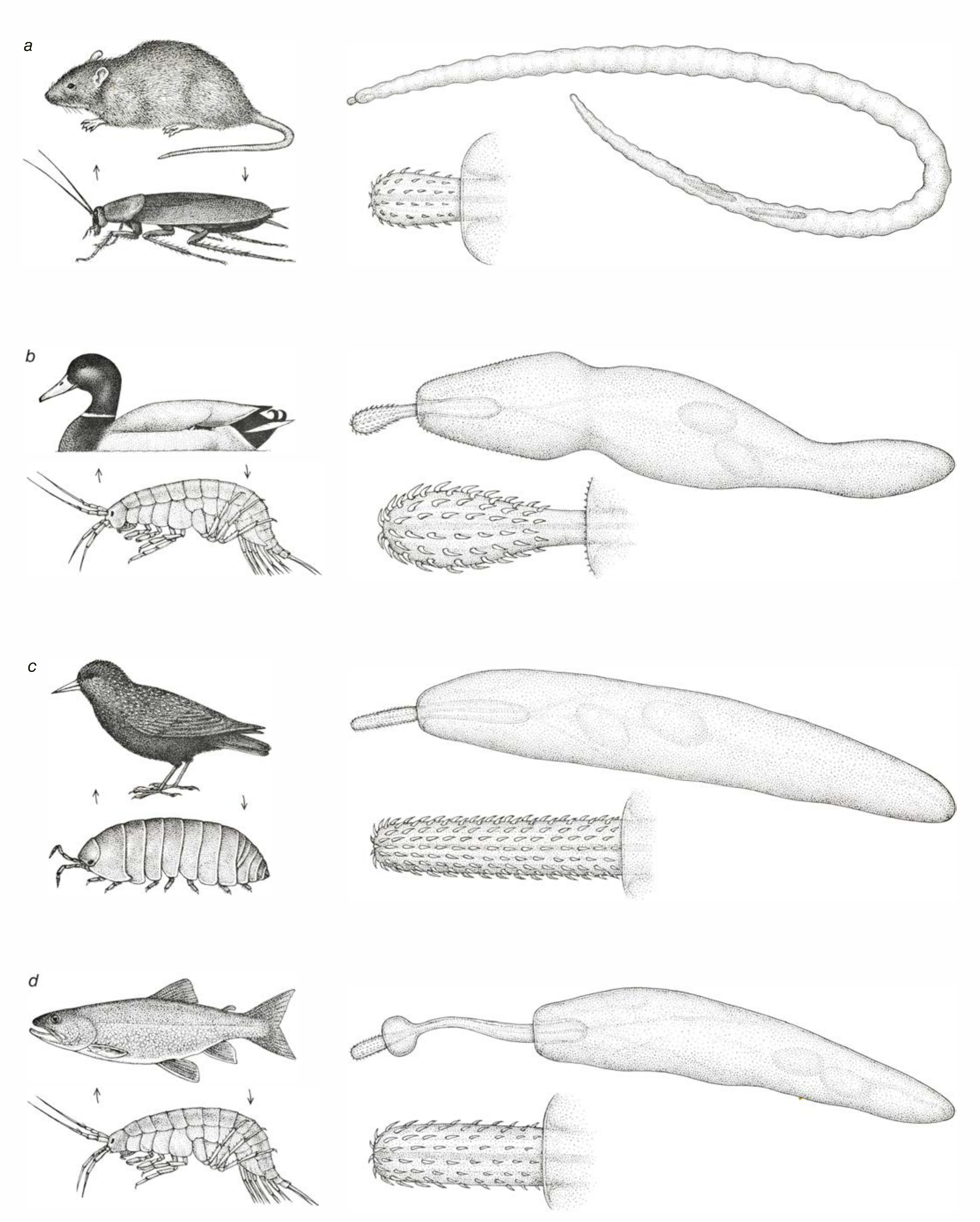
Más tarde se descubrió que algunos parásitos cambian el comportamiento de su patrón intermediario invadiendo su sistema nervioso central. El vértigo de las ovejas, una enfermedad que hace que estos rumiantes se tambaleen en círculos y se separen del rebaño, tiene su origen en la invasión del cerebro o de la médula espinal del animal por parte de la larva de la solitaria canina, *Taenia multiceps*. Los lobos y los perros salvajes, que hacen presa en estos rumiantes, son los patrones definitivos de esa solitaria.

La duela lanceolada, *Dicrocoelium dendriticum*, incide en el comportamiento de su patrón intermediario de manera más específica. *D. dendriticum* madura en las ovejas, pero pasa parte de su vida anterior en hormigas; tiene que resolver un espinoso problema: las ovejas no suelen comer hormigas. Cuando un grupo de gusanos inmaturos invade una hormiga, uno de ellos se enquista en el ganglio subesofágico, que es la parte del sistema nervioso que controla las piezas bucales y la locomoción del insecto. Wilhelm Hohorst y sus colaboradores, en Hoechst A. G., de Frankfurt, han demostrado que una hormiga infestada trepa a lo alto de una planta y, si la temperatura es lo suficientemente baja, queda sujeta a la planta mediante sus mandíbulas. De este modo aumenta la probabilidad de que se la coma una oveja que ramonea.

Al menos un grupo de parásitos induce cambios de comportamiento específicos sin dañar el tejido muscular o el nervioso de su patrón; se trata de los miembros del tipo Acantocéfalos, los llamados gusanos de cabeza espinosa. Los acantocéfalos no invaden los músculos ni el sistema nervioso, de modo que el mecanismo que subyace a sus efectos de comportamiento probablemente sea bioquímico; poco se sabe de ello, sin embargo. La capacidad que estos organismos muestran para provocar en momentos inadecuados lo que, de otro modo, serían pautas de conducta normales de sus patrones, ha atraído en los últimos años el interés de varios investigadores. Los resultados de sus estudios, incluyendo los míos propios, han demostrado que de esta capacidad están dotados los miembros de las tres clases de Acantocéfalos. Puede que sea universal en el tipo.

No se conoce con precisión el número de especies de acantocéfalos; la estimación más alta indica que el total es inferior a las 1200. Las pocas especies de las que se conoce el ciclo biológico viven como adultos en el intestino delgado de vertebrados, en particular aves y peces. Allí la hembra libera los huevos; el vertebrado los excreta y subsecuentemente los come un patrón intermediario: un artrópodo, sea un insecto o un crustáceo. El huevo hace eclosión en el intestino del artrópodo, y la larva se abre camino excavando a través de la pared intestinal hasta la cavidad del cuerpo, donde se desarrolla hasta el estadio denominado cistacanto, que puede infestar a vertebrados. Cuando un artrópodo infestado es comido por el vertebrado apropiado, el cistacanto se asienta en el intestino delgado y se desarrolla hasta convertirse en un adulto sexualmente maduro.

Desde el punto de vista morfológico, los Acantocéfalos se hallan maravillosamente adaptados a su tipo de vida parásita. El cuerpo del adulto es, en esencia, una pequeña bolsa de órganos reproductores fijada a una trompa o probóscide, cubierta ésta de ganchos curvados hacia atrás. La trompa espinosa permite al organismo fijarse a la pared intestinal del vertebrado. Los Acantocéfalos carecen de intestino pro-



1. ACANTOCEFALOS, gusanos parásitos que se ilustran aquí junto con su patrón, u hospedador, intermediario (un artrópodo), en el que se desarrolla la larva, y su patrón definitivo (un vertebrado), en la que el parásito alcanza la madurez sexual. Se ha comprobado que las cuatro especies de acantocéfalos ilustradas cambian el comportamiento de artrópodos tales como crustáceos o insectos. *Moniliformis moniliformis* (*a*) crece en cucarachas y luego infesta las ratas que se comen las cucarachas. *Polymorphus paradoxus* (*b*) se transmite a partir de anfípodos (pequeños crustáceos acuáticos) a ánades reales y otros depredadores. *Plagiorhynchus cylindraceus* (*c*), cuyo impacto sobre el comportamiento de las cochinillas de humedad fue estudiado por la autora, madura en estorninos y otros pájaros que comen cochinillas. *Pomphorhynchus laevis* (*d*) pasa de anfípodos a peces, como la trucha. La trompa espinosa con la que los gusanos adultos se fijan a la pared intestinal del vertebrado se ha dibujado a una escala común; el tamaño real de la trompa de *Plagiorhynchus* es de alrededor de un milímetro de longitud. Los propios gusanos varían mucho de tamaño: *Moniliformis* puede alcanzar hasta 30 centímetros de longitud y, las demás especies, un tamaño más de veinte veces inferior. Disponen de órganos reproductores, pero carecen de tubo digestivo y vestigio de él.

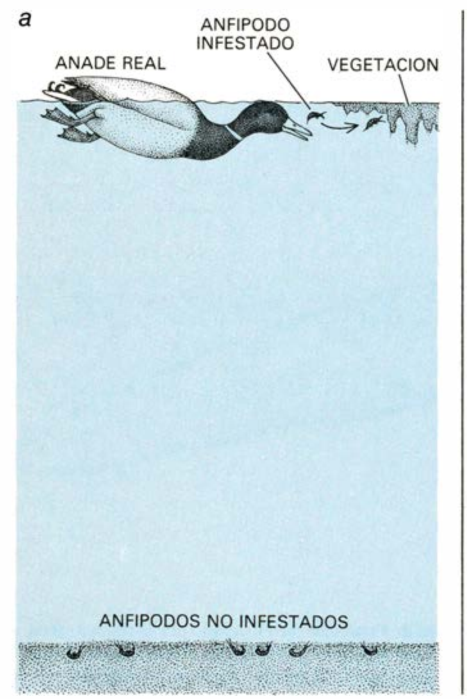
pio o de cualquier vestigio de desarrollo del mismo; pueden sobrevivir sin tubo digestivo porque habitan en un ambiente rico en nutrientes digeridos, que simplemente absorben a través de su piel. Cómo desarrollaron los Acantocéfalos estas características morfológicas y cuáles son sus parientes vivos más cercanos siguen siendo temas de especulación; y ello en buena parte porque no hay ningún otro organismo que se les parezca mucho.

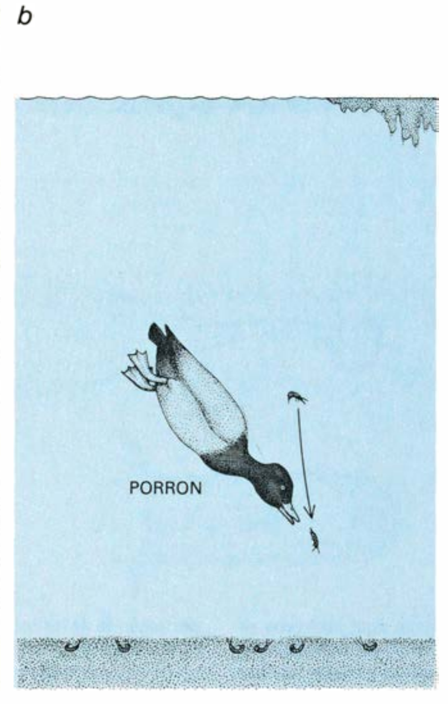
Dado el ciclo biológico del gusano, es razonable suponer que puede haber desarrollado, a través de la selección natural, rasgos que aumentan sus probabilidades de alcanzar un patrón definitivo. En la década de 1970, William M. Bethel y John C. Holmes, que trabajaban en la Universidad de Alberta en Edmonton, realizaron una serie de experimentos destinados a probar específicamente si los acantocéfalos cambiaban la respuesta de su hospedador intermediario a los estímulos ambientales. Examinaron tres especies de gusanos que tienen anfípodos (pequeños crustáceos acuáticos) como patrones intermediarios, si bien alcanzan la edad adulta en vertebrados diferentes. Encontraron que cada especie inducía distintos cambios de comportamiento que aumentaban la probabilidad de que los anfípodos fueran consumidos por el tipo apropiado de patrón definitivo.

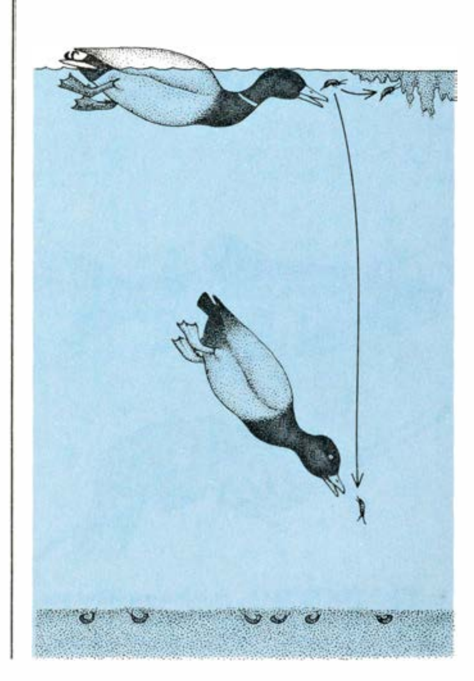
Los anfípodos no infestados huyen de la luz y rara vez se les ve en la superficie de un estanque o un lago. Cuando se les perturba, se sumergen y excavan en el fango del fondo. En contraste con éstos, Bethel y Holmes demostraron que los anfípodos infestados con cistacantos de *Polymorphus paradoxus*, que como adulto vive en el intestino delgado de ánades reales, castores y ratas almizcladas, se mueven hacia la luz. Cuando los anfípodos infestados son molestados, se deslizan por la superficie, a veces agarrándose a la vegetación o a otros objetos flotantes. Ello aumenta la probabilidad de que les coman depredadores que se alimentan en la superficie, verbigracia, ánades reales, castores y ratas almizcladas. Los crustáceos infestados con *Polymorphus marilis*, en cambio, se mueven hacia la luz pero no llegan hasta la superficie; los patrones definitivos de este parásito son patos buceadores, como los porrones bastardos. Finalmente, los anfípodos que albergan *Corynosoma constrictum* se mueven hacia la luz, aunque más de la mitad de ellos se sumergen cuando son molestados. *C. constrictum* madura tanto en patos buceadores como en patos que se alimentan en la superficie.

En pruebas efectuadas con ánades reales y ratas almizcladas en un tanque de laboratorio, Bethel y Holmes confirmaron que los cambios observados en el comportamiento afectaban a las pautas de alimentación. En comparación con los animales control no infestados, los ánades reales, que comen en superficie, tomaron un gran número de anfípodos que contenían *P. paradoxus*, un número menor (pero todavía significativo) de anfípodos infestados por *C. constrictum* y ningún anfípodo que contuviera *P. marilis*, cuyos patrones intermediarios evitan la superficie. Las ratas almizcladas comieron un número significativo de anfípodos que contenían *P. paradoxus* y ninguno con *P. marilis*. Los investigadores también encontraron que los cambios en el comportamiento de los anfípodos sólo se producían cuando los acantocéfalos habían alcanzado el estadio de cistacanto, en el que puede infestar vertebrados. Si el anfípodo fuera depredado en un estadio anterior, el parásito inmaturo no sobreviviría.

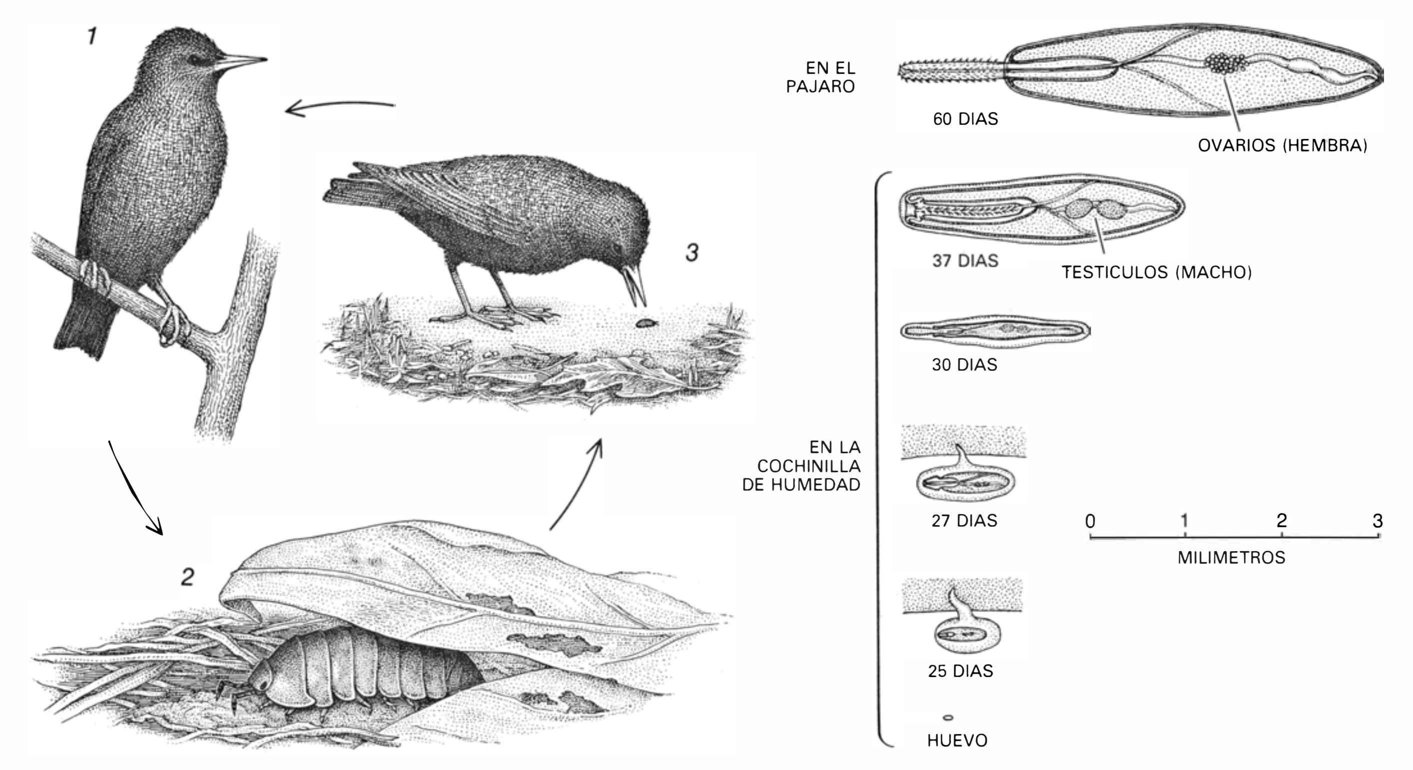
Hacia 1979 se había demostrado que siete especies de acantocéfalos, todos ellos de la clase Paleoacantocéfalos, alteraban el comportamiento de cinco especies de crustáceos acuáticos, tanto anfípodos como isópodos. Pruebas de depredación en el laboratorio sobre seis de tales parásitos habían confirmado, en cada uno de los casos, que los crustáceos infestados tenían más probabilidades de ser comidos que los controles no infestados. Me propuse extender estas investigaciones a un pa-







2. DISTINTOS IMPACTOS sobre el comportamiento de anfípodos por parte de tres especies de acantocéfalos, según demostró un experimento de laboratorio realizado por William M. Bethel y John C. Holmes, de la Universidad de Alberta. Los anfípodos no infestados evitan la luz: cuando se les molesta, excavan en el fango del fondo. Los anfípodos infestados con *Polymorphus paradoxus* (*a*) se mueven hacia la luz, y cuando se les amenaza tienden a agarrarse a la vegetación flotante o se deslizan por la superficie; allí se los comen patos nadadores, como los ánades reales. Aunque algunos crustáceos infestados con *Polymorphus marilis* (*b*) también prefieren las zonas iluminadas, no van a la superficie; son depredados por patos buceadores como los porrones bastardos. Los anfípodos que albergan *Corynosoma constrictum* (*c*) nadan hacia la superficie, pero algunos se sumergen cuando se les molesta; sus depredadores son tanto patos nadadores como patos buceadores. Los gusanos maduran en el intestino de los patos, que constituyen sus patrones definitivos.



3. CICLO BIOLOGICO DE *PLAGIORHYNCHUS CYLINDRACEUS*; comienza en el intestino delgado de un estornino, donde la hembra adulta pone huevos que luego excreta el pájaro (*1*). Cuando una cochinilla de humedad o algún otro isópodo adecuado come las heces del ave (*2*), los huevos eclosionan a las pocas horas. La larva, que inicialmente no mide más que una décima de milímetro y está armada de diminutas espinas, excava a través de la pared del tubo digestivo del isópodo y en última instancia cae a la cavidad del cuerpo, manteniéndose pegada al intestino mediante un pedúnculo. En la cavidad del cuerpo la larva aumenta considerablemente de tamaño y desarrolla los órganos del adulto. A los 60 o 65 días después de la eclosión mide ya unos cuatro milímetros y es capaz de infestar un ave. A través de un mecanismo que todavía no se conoce con exactitud, la larva altera el comportamiento de la cochinilla de humedad; por ejemplo, las cochinillas infestadas son menos propensas a guarecerse de los depredadores bajo la hojarasca. Por lo general, la trompa espinosa permanece invertida dentro del cuerpo del gusano hasta que la cochinilla es devorada por un pájaro (*3*), momento en el que la probóscide se evagina y se fija al intestino del ave. El parásito crece hasta alcanzar incluso 15 milímetros de longitud. Se cierra el ciclo con la puesta de huevos.

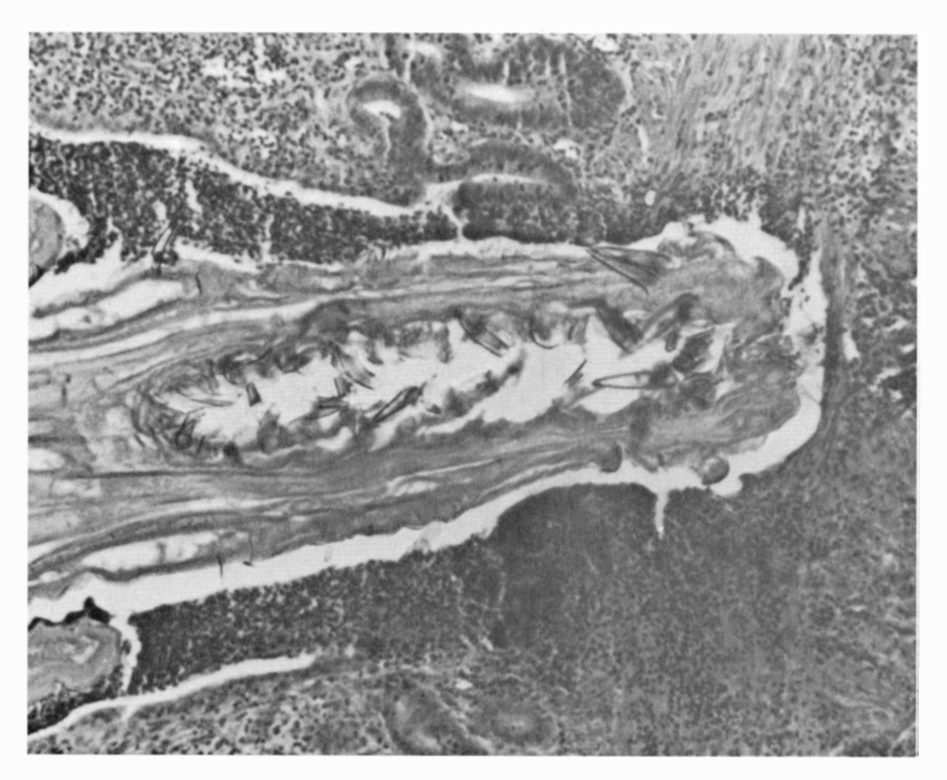
leoacantocéfalo terrestre y, por vez primera, examinar los efectos de un parásito sobre la depredación en el campo, además de en el laboratorio. Más aún: al observar el comportamiento de un acantocéfalo del que se sabía que causaba enfermedad en su patrón definitivo (ninguna de las especies estudiadas anteriormente era un patógeno conocido), esperaba determinar si el hospedador podía desarrollar un método para librarse del parásito.

Plagiorhynchus cylindraceus, citado esporádicamente y en algunos manuales como altamente patógeno para los paseriformes, parecía ser el gusano que yo andaba buscando. Cuando es adulto vive frecuentemente en estorninos, y sus patrones intermediarios son isópodos terrestres, como la cochinilla de humedad. Había buenas razones para creer que *P. cylindraceus* podía alterar el comportamiento de la cochinilla: en un isópodo de ocho milímetros su cistacanto puede alcanzar un tamaño de casi tres milímetros de longitud y uno de anchura. Además, Gerald D. Schmidt, de la Universidad de Colorado Septentrional, había señalado que las hembras de cochinilla infestadas con este parásito no desarrollaban ovarios. Finalmente, estudios de campo habían demostrado que más del 40 por ciento de los pájaros examinados estaban infestados con *P. cylindraceus* en regiones donde la presencia del gusano en los isópodos era inferior al 1 por ciento. Puesto que no se sabe que los isópodos constituyan una parte significativa de la dieta de las aves, estos resultados eran difíciles de explicar sin suponer que el acantocéfalo poseía cierta capacidad para hacer que los isópodos resultaran más atractivos para las aves.

Al objeto de examinar el comportamiento de las cochinillas de humedad en el laboratorio construí una cámara simple, a partir de dos bandejas de las que se emplean para los pasteles. La bandeja inferior, que contenía una solución salina saturada que generaba la humedad relativa deseada, estaba cubierta con una gasa de nilón sobre la que podían moverse las cochinillas. La bandeja superior, invertida, cerraba la cámara, y la unión estaba sellada mediante un burlete. Todos los experimentos se realizaron manteniendo en la cámara la misma temperatura, de 24 a 25 grados Celsius.

Di luego a comer a las cochinillas fragmentos de zanahoria recubiertos con huevos de *P. cylindraceus*. Después de esperar tres meses a que se desarrollaran los cistacantos, mezclé los animales infestados con animales control no infestados. Los experimentos se realizaron a ciegas: la infestación no cambia el color oscuro de los isópodos, de modo que yo no sabía qué animales contenían *P. cylindraceus* hasta que los disecaba después de observar su comportamiento en el aparato experimental. Con el fin de seguir el comportamiento de más de una cochinilla a la vez, las marqué con pintura de distintos colores. En una serie de experimentos examiné luego el impacto de la infestación por *P. cylindraceus* en cuatro tipos de comportamiento que quizá sean importantes a la hora de determinar si una cochinilla de humedad cae presa de un ave: la reacción del isópodo a la humedad, al refugio, a la luz y al color del sustrato (es decir, la superficie sobre la que se halla).

Comencé por ofrecer a las cochinillas la posibilidad de elegir entre una humedad relativa del 98 por ciento y otra del 75 por ciento, que generé en los extremos opuestos de un cámara dividida con diferentes soluciones salinas. Las cochinillas podían moverse libremente bajo el divisor de cristal, porque éste no llegaba a tocar la gasa de



4. TROMPA O PROBOSCIDE ESPINOSA de *P. cylindraceus*, que se aprecia clavada en el tejido intestinal de un estornino. El daño causado por el parásito suele ser local, confinado a una distancia de unos 90 micrometros de la trompa, cuyo tamaño en vivo es de alrededor de un cuarto de milímetro de anchura.

nilón. A intervalos de un minuto, durante un período aproximado de media hora, comprobaba la localización de cada isópodo. Las cochinillas infestadas se encontraban en el extremo de humedad baja con mucha más frecuencia que los controles no infestados.

De ordinario, las cochinillas de humedad se desecan y acaban muriendo cuando la humedad relativa es inferior al 98 por ciento. Puesto que las cochinillas infestadas muestran una menor aversión a la baja humedad, *P. cylindraceus* quizás aumentara su capacidad de soportar la sequedad, o bien dañara su capacidad de percibir la humedad, o ambas cosas a la vez. El mecanismo por el cual los isópodos terrestres perciben la humedad se desconoce; es imposible, pues, decir de qué modo puede alterarlo el parásito. Entra dentro de lo razonable suponer que *P. cylindraceus* logra que un isópodo soporte ambientes más secos disminuyendo la permeabilidad de su piel y reduciendo con ello su pérdida de agua, pero los resultados de experimentos que hice para comprobar esta hipótesis no aclararon nada. Hay, sin embargo, una conclusión que parece razonable: las cochinillas de humedad que frecuentan áreas secas pueden pasar más tiempo en situaciones expuestas, en las que correrán un mayor riesgo de ser depredadas por aves al acecho.

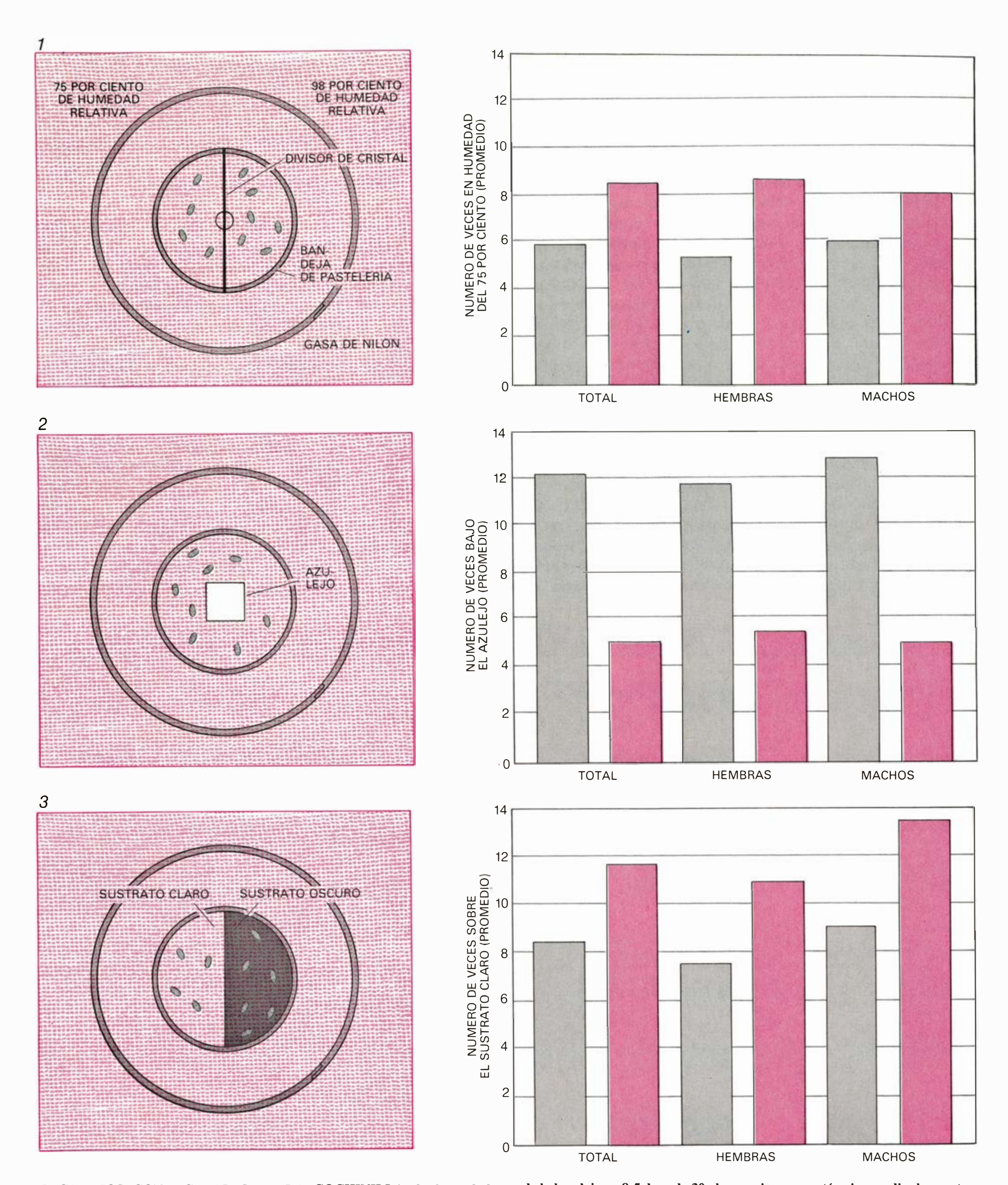
El paso siguiente fue comprobar directamente la respuesta del isópodo frente al refugio. Coloqué un azulejo sobre cuatro guijarros en el centro de la gasa de nilón y registré la localización de cada animal, de nuevo a intervalos de un minuto. El azulejo cubría el 9,5 por ciento de la superficie de la gasa de nilón, de modo que cabía esperar que una cochinilla que se moviera al azar se encontrara bajo el refugio en aproximadamente tres de cada 30 observaciones. Las cochinillas no infestadas se encontraban bajo el mosaico con mucha mayor frecuencia, lo que sugiere que de algún modo percibían y preferían el refugio; pero los animales infestados no actuaban así, lo que sugiere que ignoraban el azulejo. Tampoco aquí se conoce el mecanismo fisiológico subyacente al cambio. Resulta extraño, además, que éste no parece derivar de una respuesta diferente a la luz: cuando en un experimento distinto oscurecí la mitad de la cámara y dejé que las cochinillas se movieran con libertad, no encontré diferencia entre el comportamiento de los animales infestados y el de los que no lo estaban. Sea cual fuere el mecanismo, una respuesta aleatoria al refugio no parece ser, evidentemente, razón suficiente para que un animal evite a los depredadores.

Finalmente, probé la reacción de las cochinillas de humedad a sustratos de distintos colores. A diferencia de algunos de sus parientes acuáticos, los isópodos terrestres parasitados presentan la pigmentación normal, pero un cambio en su elección de sustrato podría destacarlos tanto como un cambio en su propio color. Una cochinilla oscura es mucho más visible sobre una acera, por ejemplo, que sobre suelo oscuro. Recreé este tipo de elección en el laboratorio, con grava de acuario, negra a un lado de la bandeja de pastel y blanca en el otro. Los resultados fueron muy claros: los isópodos infestados coincidían en la grava blanca con mucha más frecuencia que los no infestados. Puesto que un sustrato seco suele ser más claro que otro húmedo, los cambios en la respuesta a la humedad y en la elección de sustrato quizá se refuercen mutuamente.

Los experimentos de laboratorio demostraron, por tanto, que *P. cylindraceus* alteraba el comportamiento de las cochinillas de humedad, y ello aumentaba la probabilidad de que las localizaran las aves. Un isópodo infestado pasa más tiempo lejos de un refugio, sobre un sustrato claro. En el paso siguiente había que averiguar si esos cambios de comportamiento aumentaban las probabilidades de que el parásito se transmitiera a un estornino.

Con la ayuda de Danny Bell y Mike Freehling realicé estudios de campo que demuestran que así es. Hubiera sido difícil contestar a esta pregunta ciñéndonos a la observación de estorninos adultos en la naturaleza, su posterior recolección y disección. Me incliné por un grupo de sujetos experimentales más dóciles: polluelos. Preparé cajas de anidación para estorninos en el prado de la Universidad de Nuevo México. Durante la estación reproductora até limpiadores de pipa alrededor del cuello de los jóvenes estorninos por períodos de no más de dos horas, apretados lo justo para impedir que se tragaran las presas o se ahogaran. La recogida de las presas facilitaba estimar el ritmo con el que los padres proporcionaban cochinillas a sus crías (alrededor de una cochinilla cada 10 horas); pude comprobar, además, que los polluelos no adquirían ningún gusano de *P. cylindraceus* de presas que no fueran cochinillas de humedad. Al mismo tiempo recolecté 250 cochinillas de los alrededores de las cajas de anidación, donde los estorninos adultos buscaban la comida. A partir de esa muestra determiné que aproximadamente el 0,4 por ciento de la población de cochinillas de la zona albergaba larvas de *P. cylindraceus*.

Sin embargo, cuando disequé 22 po-



5. CAMBIOS CONDUCTUALES DE LA COCHINILLA de humedad, según observaciones de la autora tras comparar cochinillas infestadas con *P. cylindraceus* (*color*) con controles no infestados (*gris*) en una serie de experimentos de laboratorio. La "cámara de comportamiento" estaba compuesta por una bandeja de las de pastelería con una gasa de nilón extendida sobre ella; una segunda bandeja invertida se disponía sobre la primera. Después de un período de aclimatación, grupos de diez cochinillas marcadas individualmente podían moverse libres por la gasa de nilón durante media hora. La situación de cada isópodo se registraba a intervalos de un minuto. Primero se permitió que las cochinillas eligieran entre humedades relativas del 98 y del 75 por ciento (*1*). Los isópodos infestados se hallaban en el lado de humedad baja de la bandeja en 8,5 de cada 30 observaciones, por término medio; los controles no infestados se hallaban en esa parte en menos de 6 observaciones. Una diferencia aún más sorprendente se encontró en la respuesta de los isópodos a un azulejo elevado que simulaba un refugio (*2*): las cochinillas no infestadas se hallaban bajo el azulejo, que cubría el 9,5 por ciento de la gasa, en más de 12 de cada 30 observaciones, mientras que las cochinillas infestadas se hallaban bajo el refugio con una frecuencia sólo ligeramente superior a la que se esperaría de un movimiento al azar. Finalmente, los isópodos infestados coincidían con más frecuencia sobre un sustrato claro (*3*), del que destacaban considerablemente. Las cochinillas de humedad que se desplazan por terrenos abiertos o sobre sustratos claros tienen más probabilidades de ser devoradas.

lluelos, encontré al menos un *P. cylindraceus* en siete de ellos, esto es, el 32 por ciento. Dada la proporción del parásito en la población de cochinillas de humedad, la tasa de administración de cochinillas y el hecho de que ninguno de los polluelos tuviera más de 18 días de edad, sólo alrededor de la mitad de este número de crías debiera haber contenido gusanos si los padres obtenían su alimento al azar, es decir, si los cambios de comportamiento no estaban haciendo que las cochinillas infestadas fueran presas más atractivas. Además, el ritmo de suministro de cochinillas era, en la mayoría de los casos, de menos de una cada 10 horas, y la mayoría de polluelos tenía menos de 18 días de edad. Si en lugar de los máximos hubiera utilizado las medias de esos factores, la tasa estimada de transmisión del parásito para una alimentación aleatoria hubiera sido incluso menor. La tasa de transmisión real hubiera sido mayor si no hubiera contado como infestaciones simples aquellos casos en los que un polluelo presentaba más de un gusano. Así, pues, la estimación anterior de la diferencia entre la tasa de transmisión observada y la tasa atribuida a alimentación aleatoria es conservadora; por su parte, los indicios de que *P. cylindraceus* influye en la depredación de las cochinillas por los estorninos son, probablemente, mayores de lo que acabo de sugerir.

A resultados similares se llegó en tres estudios de campo subsiguientes en los que disequé un total de 96 polluelos y 3000 cochinillas de humedad. Para afianzar esos indicios, y para observar más directamente la respuesta de los estorninos al comportamiento alterado de las cochinillas de humedad infestadas, realicé asimismo una prueba de depredación en el laboratorio. Recubrí la mitad del fondo de un contenedor con arena negra, húmeda, y la otra mitad con arena blanca, más seca. Luego puse en el contenedor el mismo número de cochinillas infestadas y no infestadas. Les dejé plena libertad de movimiento por la noche; expuse el contenedor a un estornino cautivo y lo observé hasta que se comió casi la mitad de las cochinillas. Los isópodos negros resaltaban sobre la arena blanca; en las repeticiones del experimento los pájaros tomaron siempre la mayor parte de sus presas de esa zona. Así, los estorninos ingirieron un número significativamente mayor de cochinillas infestadas que de no infestadas.

La habilidad que *P. cylindraceus* muestra al conseguir aumentar la probabilidad de ser transportado desde una cochinilla de humedad hasta un pájaro plantea una interesante cuestión: puesto que se supone que el parásito es patógeno, ¿por qué razón un estornino come cochinillas infestadas y las ofrece a sus crías? Quizás el ave no pueda distinguir las cochinillas infestadas de las que no lo están (los dos tipos de animales parecen iguales a los ojos humanos). Pero, entonces, cabe preguntarse por qué un estornino no evita por completo a las cochinillas, por lo demás una parte insignificante de su dieta. La respuesta, creo yo, hay que buscarla en la inocuidad de *P. cylindraceus*, a pesar de que comporte efectos

ESPECIES PARASITAS	PATRON INTERMEDIARIO	CAMBIO DE COMPORTAMIENTO	PATRON DEFINITIVO
ARQUIACANTOCEFALOS			
MONILIFORMIS MONILIFORMIS	*PERIPLANETA AMERICANA* (CUCARACHA)	HIPERACTIVA: SE DESPLAZA HACIA LA LUZ	RATA
PALEOACANTOCEFALOS			
POLYMORPHUS PARADOXUS	*GAMMARUS LACUSTRIS* (ANFIPODO)	SE DESPLAZA HACIA LA LUZ; SE SUJETA A LA VEGETACION EN LA SUPERFICIE AUNQUE SE LE MOLESTE	ANIMALES QUE SE ALIMENTAN EN LA SUPERFICIE: ANADE REAL, RATA ALMIZCLADA, CASTOR
POLYMORPHUS MARILIS	*GAMMARUS LACUSTRIS*	PREFIERE ZONAS ILUMINADAS PERO NO VA A LA SUPERFICIE	PATOS BUCEADORES: PORRONES
CORYNOSOMA CONSTRICTUM	*HYALELLA AZTECA* (ANFIPODO)	SE DESPLAZA HACIA LA LUZ Y VA A LA SUPERFICIE; ALGUNOS ANIMALES NO SE HUNDEN CUANDO SE LES PERTURBA	ANIMALES QUE SE ALIMENTAN EN LA SUPERFICIE Y PATOS BUCEADORES: ANADE REAL Y PORRON
POLYMORPHUS MINUTUS	*GAMMARUS LACUSTRIS*	SE DESPLAZA HACIA LA LUZ	PATOS QUE COMEN EN SUPERFICIE
PLAGIORHYNCHUS CYLINDRACEUS	*ARMADILLIDIUM VULGARE* (COCHINILLA DE HUMEDAD)	PASA MAS TIEMPO EN ZONAS DE MENOR HUMEDAD, LEJOS DE UN REFUGIO Y SOBRE UN SUSTRATO MAS CLARO	PASERIFORMES
POMPHORHYNCHUS LAEVIS	*GAMMARUS PULEX* (ANFIPODO)	PASA MAS TIEMPO EN AGUAS ABIERTAS; VA A LA LUZ Y REPOSA EN LA VEGETACION SUPERFICIAL; NADA EN ESPIRAL	PECES: TRUCHAS
ACANTHOCEPHALUS DIRUS	*ASELLUS INTERMEDIUS* (ISOPODO ACUATICO)	MAS ACTIVO	PECES: *SEMOTILUS ATROMACULATUS*
ACANTHOCEPHALUS JACKSONI	*LIRCEUS LINEATUS* (ISOPODO ACUATICO)	MAS ACTIVO: PREFIERE SUPERAR LOS OBSTACULOS POR ENCIMA QUE POR DEBAJO	PECES
EOACANTOCEFALOS			
NEOECHINORHYNCHUS CYLINDRATUS	*PHYSOCYPRIA PUSTULOSA* (OSTRACODO)	VA HACIA LA SUPERFICIE	PECES: *GAMBUSIA*
OCTOSPINIFEROIDES CHANDLERI	*CYPRIDOPSIS VIDUA* Y *PHYSOCYPRIA PUSTULOSA* (OSTRACODOS)	VA HACIA LA SUPERFICIE; PREFIERE ZONAS ILUMINADAS	PECES

6. ONCE ESPECIES DE ACANTOCEFALOS, que comprenden miembros de las tres clases del tipo, alteran el comportamiento de sus hospedadores intermediarios. En cada caso, el cambio parece determinar que el patrón intermediario se convierta en una presa más fácil para el patrón definitivo.

adversos sobre algunas aves y de lo que indiquen algunos informes al respecto.

La pérdida de peso es un signo común de mala salud, pero no encontré relación alguna entre el peso y la presencia o el número de gusanos de *P. cylindraceus*, ni en los polluelos que examiné ni en 103 estorninos adultos que atrapé. Además, cuando se examina al microscopio los tejidos de un ave infestada, no hay síntoma de daño más allá de los 90 micrometros de la trompa del acantocéfalo. Si el parásito sólo raramente provoca enfermedad, o si sus efectos adversos son, por lo general, mínimos, quizá no le salga a cuenta al estornino privarse de las cochinillas. Si a ello añadimos que la alteración del comportamiento hace de las infestadas presas más fáciles, tendremos la razón de la presencia de *P. cylindraceus* en tantos estorninos.

Más alto es el precio que paga la cochinilla por albergar a *P. cylindraceus*. Aumenta las probabilidades de ser comida y, si es hembra, no desarrolla ovarios, ni, por tanto, se reproduce. La baja incidencia del parásito en las poblaciones naturales de cochinillas quizá dé a entender que los isópodos evitan el parásito o, simplemente, puede reflejar que las cochinillas infestadas sufren más depredación. Para determinar si las cochinillas, en especial las hembras, hacen algún esfuerzo por evitar consumir huevos de acantocéfalo, sequé y pesé heces de aves y ofrecí separadamente dos porciones a 26 hembras distintas. Una parte estaba recubierta con una suspensión acuosa de huevos de *P. cylindraceus*; la otra, sólo se había rehidratado. Transcurrida una semana, sequé y pesé los fragmentos restantes. Las cochinillas de humedad mostraron una tendencia a "evitar" las heces "infestadas", pero el resultado no fue estadísticamente significativo y, en cualquier caso, la tendencia no parecía espectacular, dados los riesgos implicados. Por ser omnívoras, quizá las cochinillas de humedad no discriminen demasiado. También pudiera ser que los beneficios de comer heces de aves, ricas en nutrientes, superen al riesgo de ser devoradas por las aves.

La capacidad de alterar el comportamiento del patrón no es exclusiva de los paleoacantocéfalos, como *P. cylindraceus*, sino que se extiende a las otras dos clases de acantocéfalos. David J. DeMont y Kenneth C. Corkum, de la Universidad estatal de Louisiana, descubrieron recientemente tal capacidad en dos eoacantocéfalos, que maduran en peces y tienen como patrones intermediarios a ostrácodos (otro tipo de crustáceos acuáticos). Los dos parásitos problablemente incrementan la vulnerabilidad de los ostrácodos intensificando su atracción por la luz.

A las cucarachas infestadas con *Moniliformis moniliformis*, componente de la tercera clase de acantocéfalos (Arquiacantocéfalos), también les atrae la luz. Además, son hiperactivas, como descubrí cuando coloqué individuos infestados en ruedas de actividad conectadas a aparatos que registraban la frecuencia del movimiento. Las cucarachas hiperactivas que se desplazan hacia la luz tienen menos probabilidades de permanecer escondidas y facilitan, por contra, que se las coman las ratas, patrones definitivos de *M. moniliformis*.

Los efectos que estos acantocéfalos y otros tienen sobre sus patrones no son tan mágicos como podría parecer. Sobre todo porque los parásitos no inducen pautas de comportamiento nuevas, sino que, simplemente, desencadenan pautas ya existentes en momentos fatalmente inoportunos. Lo que, por otro lado, no deja de constituir una hazaña; queda todavía por elaborar una explicación fisiológica general al modo en que un acantocéfalo la logra mientras se mantiene flotando en la cavidad del cuerpo del patrón.

La constatación de que los parásitos pueden cambiar el comportamiento del patrón plantea muchas incógnitas sugestivas. Los zoólogos de campo deben caer en la cuenta de la posibilidad de que el comportamiento observado esté "manipulado". Un parásito puede influir incluso sobre la evolución de su patrón, igual que muchos depredadores influyen en la evolución de su presa. Aunque este fenómeno no es evidente en el caso de *P. cylindraceus*, un parásito verdaderamente patógeno podría hacer que el hospedador desarrollara sistemas para evitar al parásito o para resistir sus efectos patógenos; cualquier animal con estas capacidades tendría ventaja sobre otros miembros de su especie. William D. Hamilton y Marlene Zuk, de la Universidad de Michigan, acaban de sugerir que la resistencia a las enfermedades parasitarias se expresa en el aspecto físico y en la exhibición de cortejo de los paseriformes, y puede afectar a la elección de la pareja. Se precisa más información para confirmar tal relación aunque, a partir de estas investigaciones y de otras semejantes, resulta claro que los biólogos están empezando a comprender el grado en que los parásitos pueden influir en el comportamiento de otros organismos.

El universo inflacionario

De acuerdo con una nueva teoría cosmológica, el universo estaría inmerso en una región del espacio mucho más amplia. Esta región registró un crecimiento extraordinario una fracción de segundo después de la gran explosión inicial

Alan H. Guth y Paul J. Steinhardt

En estos últimos años, cabos sueltos de la teoría estándar de la gran explosión han instado el desarrollo de un nuevo modelo sobre la historia de los primeros momentos del universo. Este modelo recibe el nombre de universo inflacionario; concuerda exactamente con la descripción aceptada del universo observado para tiempos posteriores a los primeros 10^{-30} segundos. Mas, para esta primera fracción de segundo, la explicación cambia por completo. Según el nuevo modelo, el universo tuvo un breve período de inflación, o expansión, extraordinariamente rápida en la que su diámetro creció en un factor quizás 10^{50} veces mayor que el que se ha venido admitiendo. En ese momento de enorme crecimiento, podría haberse creado toda la materia y energía del universo, virtualmente a partir de la nada. El proceso inflacionario comporta también implicaciones importantes para nuestro universo. Si el nuevo modelo es correcto, el universo observado se limita a una porción muy pequeña de cuanto es en su integridad.

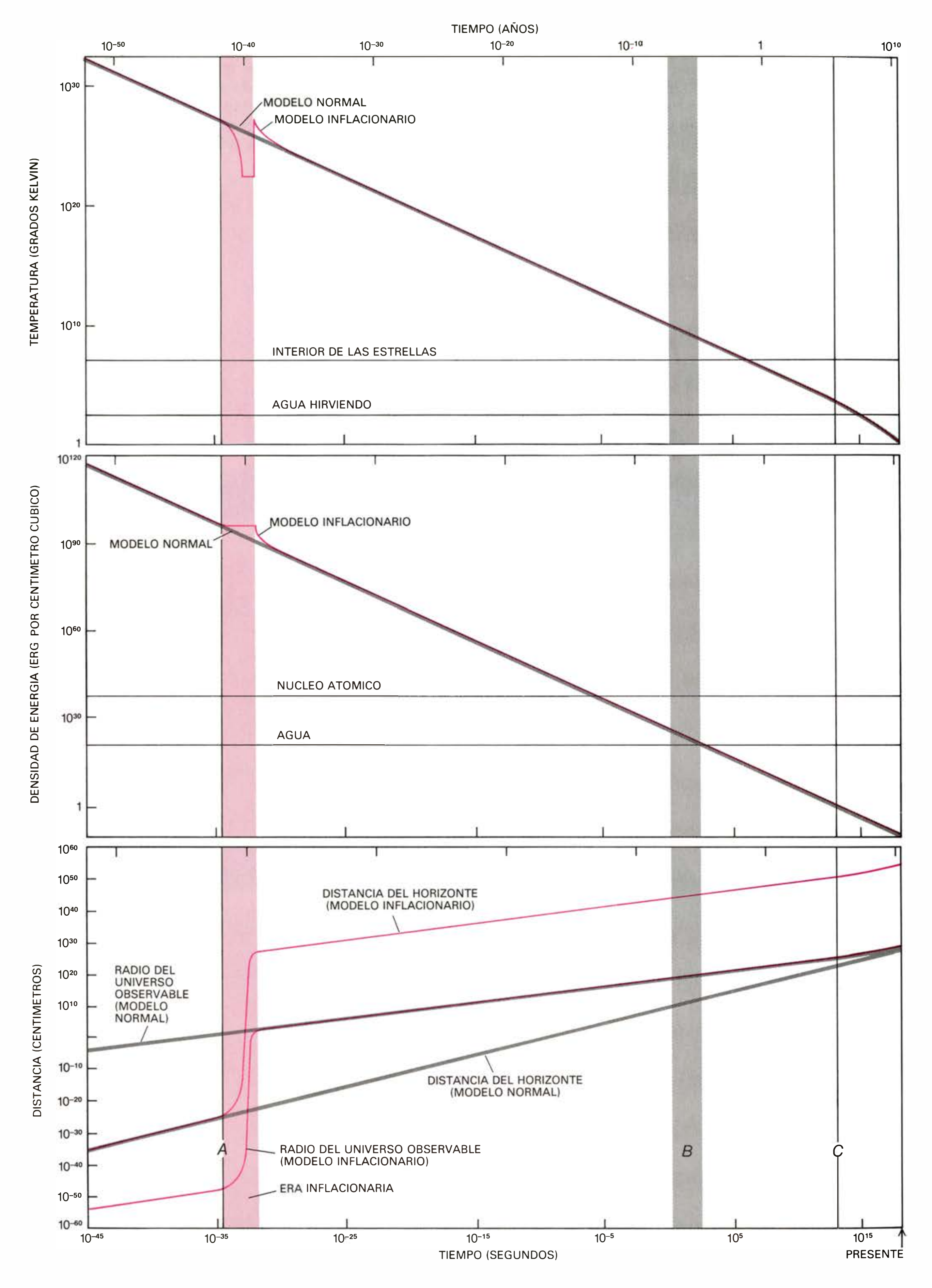
Muchas características del modelo inflacionario las comparte el modelo estándar de la gran explosión. Para ambos, el comienzo del universo se sitúa a unos 10.000 o 15.000 millones de años, partiendo de una bola de fuego primitiva, con densidad y temperatura extremas; desde entonces se está expandiendo y enfriando. Esta descripción ha permitido justificar muchos aspectos del universo observado; entre ellos, el corrimiento hacia el rojo de la luz procedente de las galaxias lejanas, la radiación cósmica de fondo de microondas y la abundancia primitiva de elementos ligeros. Todas estas predicciones están relacionadas con sucesos que seguramente ocurrieron después del primer segundo, cuando ya los dos modelos coinciden.

Hasta hace unos cinco años, eran pocos los intentos serios para describir cómo era el universo del primer segundo. Se creía que la temperatura en este período debía superar los 10.000 millones de grados Kelvin y apenas si se sabía algo de las propiedades que adornan a la materia en tales condiciones. A pesar de tan escaso bagaje, y apoyándose en los últimos resultados conseguidos en la física de partículas elementales, los cosmólogos se han empeñado en desentrañar la historia del universo hasta retroceder a los 10^{-45} segundos de edad. (En el intervalo anterior, la densidad de energía debió ser de tal magnitud que habría que sustituir la teoría de la relatividad general de Einstein por una teoría cuántica de la gravedad que, por ahora, no existe.) Cuando se pretende extender a estos momentos iniciales el modelo normal de la gran explosión surgen varios problemas. En primer lugar, se pone de manifiesto que el modelo requiere bastantes hipótesis, rigurosas e inexplicadas, sobre las condiciones iniciales del universo. Y, en otro sentido, la mayoría de las nuevas teorías de partículas elementales revelan que el modelo estándar conduciría a una hiperproducción de partículas exóticas llamadas monopolos magnéticos. (Un monopolo es un polo magnético, norte o sur, aislado.)

El modelo inflacionario se ideó para resolver estos problemas. Las ecuaciones que describen el período infacionario presentan una característica muy atractiva: a partir de casi todas las condiciones iniciales, el universo evoluciona hacia el estado que, por hipótesis, debe tomarse como inicial en el modelo normal o estándar. Además, la densidad de monopolos magnéticos que predice, pequeña, resulta compatible con las observaciones. Dentro del marco que nos prestan los recientes avances en la teoría de partículas elementales, el modelo inflacionario llega como una solución natural de muchos problemas que presenta la descripción normal de la gran explosión.

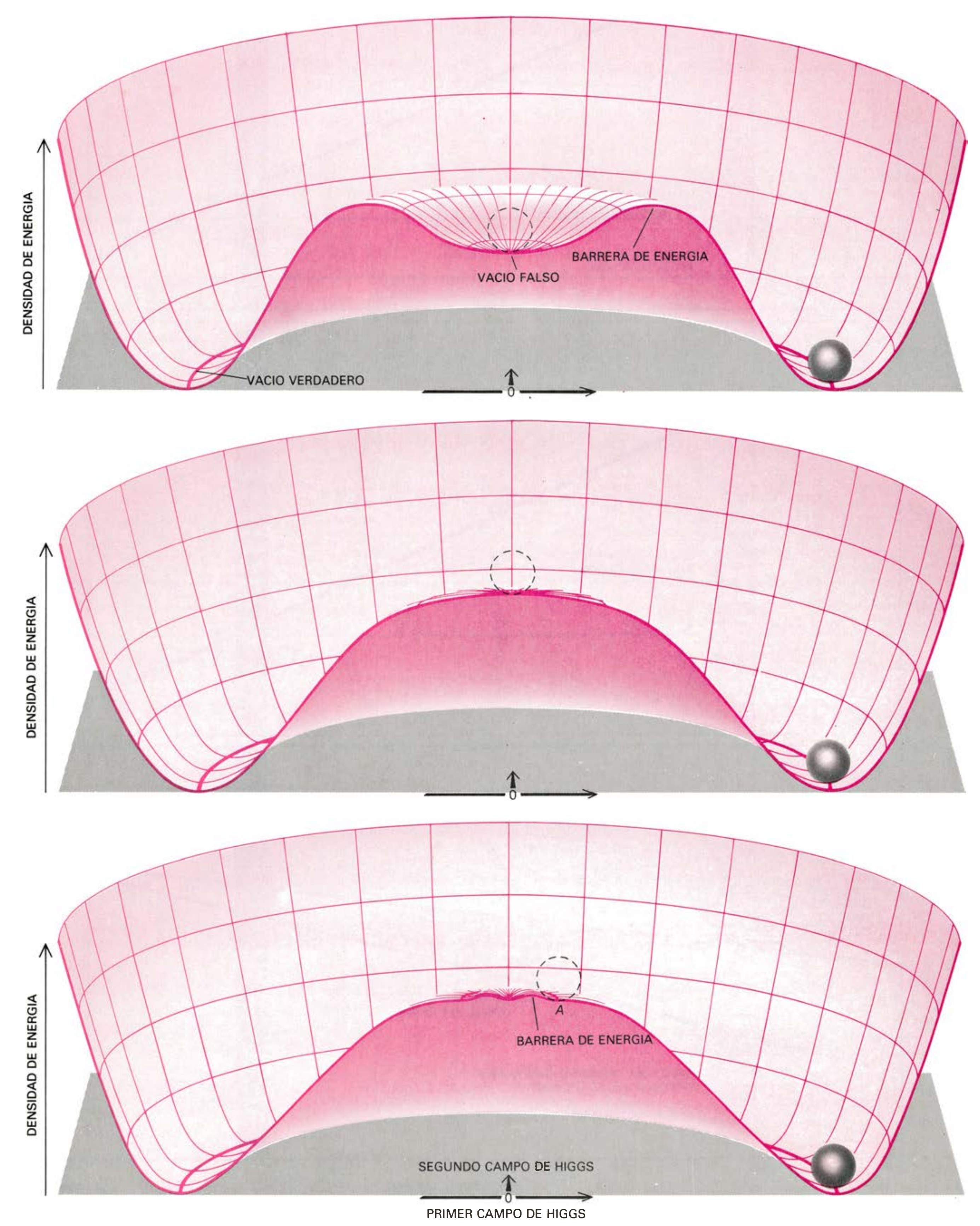
El modelo normal de la gran explosión se basa en varias hipótesis. Primera: suponer que las leyes fundamenta-

1. MODELO INFLACIONARIO del universo, representado en este conjunto de gráficos por las curvas coloreadas, que indican cómo pueden haber evolucionado varias propiedades del universo observado, a partir de los 10^{-45} segundos después de la gran explosión inicial. Las curvas grises representan el modelo normal de la gran explosión, modelo que coincide con el inflacionario para todos aquellos tiempos posteriores a los 10^{-30} segundos. A efectos comparativos, la gráfica de temperatura (*arriba*) incluye el punto de ebullición del agua (373 grados Kelvin) y la temperatura en el centro de una estrella típica (10 millones de grados Kelvin). Análogamente, la gráfica de la densidad de energía (*en medio*) señala la densidad de energía del agua (10^{21} erg por centímetro cúbico) y la de un núcleo atómico (10^{36} erg por centímetro cúbico). En la gráfica de las dimensiones espaciales (*abajo*), cada uno de los modelos cosmológicos está representado por dos curvas. Una curva recoge la evolución de aquella región del espacio que resultará ser el universo observado, mientras que la otra indica la distancia del horizonte: la distancia total que debe haber recorrido una señal de luz desde el inicio del universo. El modelo normal presenta el llamado problema del horizonte, que se plantea porque la distancia del horizonte, durante la mayor parte de su historia, es mucho menor que el radio del universo observable. En el modelo inflacionario, la distancia del horizonte, en cualquier tiempo, es mucho mayor que el radio del universo observable. En el eje temporal están indicados varios sucesos significativos. *A* designa el instante en que tiene lugar la transición de la fase en el modelo estándar (normal) de la gran explosión, deducido de las teorías de gran unificación de las interacciones de partículas elementales; para tiempos anteriores, se cree que las altas temperaturas dominantes hacen que las distintas fuerzas no gravitacionales que actúan entre las partículas estén relacionadas entre sí por una simetría que se rompió espontáneamente cuando la temperatura descendió por debajo del valor crítico de unos 10^{27} grados. Rasgo fundamental del modelo inflacionario es la prolongación de la transición de fase, que se extiende a lo largo de un período llamado era inflacionaria (*banda coloreada*); durante esta época, el universo se expande multiplicando su tamaño por un factor enorme, tal vez 10^{50}, o mayor. Simultáneamente, desciende la temperatura; sin embargo, por efectos cuánticos que aparecen en el contexto de la relatividad general, se estabiliza alrededor de los 10^{22} grados. La banda gris, *B*, indica el período durante el cual se formaron la mayoría de los núcleos atómicos más ligeros: *C* indica el instante a partir del cual el universo se volvió transparente para la radiación electromagnética.



TIEMPO (AÑOS)
10⁻⁵⁰
10⁻⁴⁰
10⁻³⁰
10⁻²⁰
10⁻¹⁰
1
10¹⁰
TEMPERATURA (GRADOS KELVIN)
10³⁰
10²⁰
10¹⁰
1
MODELO NORMAL
MODELO INFLACIONARIO
INTERIOR DE LAS ESTRELLAS
AGUA HIRVIENDO
DENSIDAD DE ENERGIA (ERG POR CENTIMETRO CUBICO)
10¹²⁰
10⁹⁰
10⁶⁰
10³⁰
1
MODELO INFLACIONARIO
MODELO NORMAL
NUCLEO ATOMICO
AGUA
DISTANCIA (CENTIMETROS)
10⁶⁰
10⁵⁰
10⁴⁰
10³⁰
10²⁰
10¹⁰
10⁻¹⁰
10⁻²⁰
10⁻³⁰
10⁻⁴⁰
10⁻⁵⁰
10⁻⁶⁰
DISTANCIA DEL HORIZONTE (MODELO INFLACIONARIO)
RADIO DEL UNIVERSO OBSERVABLE (MODELO NORMAL)
DISTANCIA DEL HORIZONTE (MODELO NORMAL)
A
B
C
RADIO DEL UNIVERSO OBSERVABLE (MODELO INFLACIONARIO)
ERA INFLACIONARIA
10⁻⁴⁵
10⁻³⁵
10⁻²⁵
10⁻¹⁵
10⁻⁵
10⁵
10¹⁵
TIEMPO (SEGUNDOS)
PRESENTE

les de la física no cambian con el tiempo y que los efectos gravitacionales quedan correctamente descritos por la teoría de la relatividad general de Einstein. Segunda: admitir que el universo primitivo estaba ocupado por un gas muy caliente de partículas elementales, casi uniforme, en expansión y en equilibrio térmico. Este gas llenaba todo el espacio, y ambos se expandían conjuntamente con la misma velocidad. Al promediarlas en grandes regiones, las densidades de materia y energía permanecían prácticamente uniformes de un sitio a otro, mientras proseguía su evolución el universo. Tercera hipótesis: suponer que los cambios ocurridos en el estado de la materia o la radiación fueron tan suaves que tuvieron un efecto despreciable en la historia termodinámica del universo. La violación de



esta última hipótesis es la clave del modelo inflacionario del universo.

El modelo de la gran explosión da lugar a tres importantes predicciones que pueden comprobarse. De acuerdo con la primera, a medida que el universo se expande, las galaxias se alejan unas de otras con una velocidad proporcional a la distancia que las separa. Fue en los años veinte cuando Edwin P. Hubble dedujo esta ley de la expansión, fundándose en sus observaciones del corrimiento hacia el rojo de la luz procedente de las galaxias lejanas. El modelo de la gran explosión predice, en segundo lugar, que, bañando el universo, debería haber una radiación de fondo de microondas, resto del intenso calor que reinaba en sus orígenes. Centenares de miles de años después de la gran explosión, el universo se volvió transparente para esta radiación. Desde entonces, la materia se ha ido agrupando en estrellas, galaxias y todo lo demás, mientras que la radiación ha continuado simplemente expandiéndose y corriéndose hacia el rojo, en definitiva, enfriándose. En 1964, Arno A. Penzias y Robert W. Wilson, de los Laboratorios Bell, descubrieron una radiación de fondo de microondas que se recibía uniformemente de todas las direcciones, con una temperatura eficaz de unos tres grados Kelvin. Y vayamos con la tercera predicción: la relativa a la formación de núcleos atómicos ligeros, a partir de protones y neutrones, durante los tres primeros minutos subsiguientes a la gran explosión. De la teoría se deducen valores aceptables para la proporción de helio-4, deuterio, helio-3 y litio-7. (Se cree que los núcleos pesados se produjeron mucho más tarde, en el interior de las estrellas.)

A pesar de los éxitos del modelo de la gran explosión, todos ellos relacionados con sucesos que acontecieron cuando el universo tenía ya un segundo o más de vida, existen dificultades asociadas al estado del universo mucho antes de que cumpliera el primer segundo. Un grupo de esas dificultades guarda relación con las condiciones especiales que exige el modelo al estado inicial del universo surgido de la gran explosión.

El primer problema está en explicar la uniformidad a gran escala del universo observado, uniformidad que se manifiesta con mayor evidencia en la radiación de fondo de microondas; presenta ésta una temperatura uniforme en una parte sobre 10.000. En el modelo normal, el universo evoluciona con excesiva rapidez y no permite que esta uniformidad se logre a través de los procesos habituales por los que un sistema alcanza su equilibrio térmico. El motivo está en que ninguna información o proceso físico puede propagarse más deprisa que la señal luminosa. Para cualquier tiempo dado, existe una distancia máxima, conocida por distancia del horizonte, que nos indica el límite hasta donde puede haber llegado una señal luminosa desde el principio del universo. En el modelo estándar, las fuentes de la radiación de fondo de microondas correspondientes a observaciones procedentes de regiones opuestas del firmamento estaban separadas por una distancia más de noventa veces superior a la distancia del horizonte cuando se emitió la radiación. Ahora bien, estas regiones no podían comunicarse: ¿cómo explicar su evolución hacia estados prácticamente tan idénticos?

2. DENSIDAD DE ENERGIA del universo en función de dos campos de Higgs. Eso representan los diagramas tridimensionales de la página opuesta. Los campos pertenecen a un conjunto especial, postulado en las teorías de gran unificación para explicar como se rompe espontáneamente la simetría. Las superficies, de la que se muestra una sección transversal, tienen simetría rotacional alrededor del eje vertical, que corresponde a un estado en que ambos campos de Higgs son nulos. Sin contar con las perturbaciones térmicas, este estado de simetría no rota, conocido por falso vacío, tendría una densidad de energía de unos 10^{95} erg por centímetro cúbico, o unas 10^{59} veces superior a la densidad de energía de un núcleo atómico. La simetría rotacional se rompe cuando uno de los campos de Higgs (o los dos) toma valores no nulos. En esta ocasión, la teoría se ha formulado de suerte que los estados con la densidad de energía más baja, conocidos por estados de vacío verdadero, son estados con la simetría rota y forman un círculo en el plano horizontal, al pie de cada uno de los diagramas. La evolución del universo, en esta analogía, puede describirse imaginando una bola que rueda por la superficie. La distancia de la bola al eje central indica los valores combinados de los campos de Higgs; la altura respecto de la superficie horizontal, la densidad de energía del universo. Cuando los dos campos de Higgs son nulos, la bola se halla en el eje de simetría; cuando los campos de Higgs tienen un valor que corresponde a la más baja de las posibles densidades de energía, la bola está situada en alguna parte del foso que define los estados de simetría rota o vacío verdadero. En la forma original del modelo de universo inflacionario, se suponía que la función densidad de energía tenía la forma indicada en el diagrama superior. El fenómeno de la inflación habría ocurrido mientras el universo se encontraba en el estado de falso vacío. Si fuesen aplicables las leyes de la física clásica, este estado sería totalmente estable, pues no existiría ninguna energía disponible para conseguir que los campos de Higgs atravesaran la barrera energética. Sin embargo, según las leyes de la física cuántica, los campos en pequeñas regiones del espacio pueden, por "efecto túnel", atravesar la barrera de energía formando burbujas de la fase de simetría rota que, enseguida, empiezan a crecer. En el nuevo modelo inflacionario (*diagrama central*) no hay ninguna barrera de energía; en su lugar, el vacío falso está en la cima de una meseta casi llana. En estas circunstancias, la transición desde el falso vacío hasta la fase de simetría rota se produce a través de un mecanismo de suave rodamiento: los campos de Higgs se separan de su valor nulo inicial debido a fluctuaciones térmicas o cuánticas, y tienden hacia sus valores del vacío verdadero tal como una bola se deslizaría por una meseta del mismo perfil. La expansión acelerada del universo tiene lugar durante los primeros momentos de la caída en rotación, mientras la densidad de energía permanece prácticamente constante. Un solo dominio de la fase de simetría rota podría crecer, entonces, lo suficiente para abarcar todo el universo observable. Cuando los campos de Higgs llegan al fondo del foso, oscilan alrededor del mínimo de la densidad de energía, provocando un recalentamiento del universo. En una variante del nuevo modelo inflacionario (*diagrama inferior*), el falso vacío queda rodeado por una pequeña barrera de energía. Como en el modelo inflacionario original, el falso vacío se desintegra con la formación aleatoria de burbujas, creadas por efecto túnel de los campos de Higgs a través de la barrera de energía. Como en este caso la barrera de energía es pequeña, por efecto túnel los campos de Higgs sólo llegan hasta el círculo indicado con la letra *A*. Ahora bien, la pendiente en *A* es casi llana; en consecuencia, los campos de Higgs evolucionan muy lentamente hasta tomar los valores del vacío verdadero. La expansión acelerada del universo tiene lugar mientras los campos de Higgs están cerca de *A*; y una sola burbuja puede crecer lo suficiente para abarcar todo el universo observable.

Se llama problema del horizonte a la dificultad que ofrece justificar la uniformidad del universo sobre distancias grandes comparadas con las del horizonte. No se trata de ninguna contradicción del modelo estándar; si incluimos dicha uniformidad en las condiciones iniciales, el universo evolucionará de manera uniforme. El problema estriba en que uno de los hechos más sobresalientes del universo observado, su uniformidad a gran escala, no halla explicación en el marco del modelo normal; debe incluirse como condición inicial.

Pero no basta la hipótesis de la uniformidad a gran escala; si el modelo normal de la gran explosión pretende explicar la no uniformidad observada a escalas más pequeñas, ha de echar mano de otra hipótesis más. Para que se produzca el agrupamiento de la materia en galaxias, cúmulos de galaxias, supercúmulos de cúmulos, y así sucesivamente, debemos incluir entre las condiciones iniciales un espectro original de inhomogeneidades. El hecho de que el espectro de inhomogeneidades no pueda explicarse constituye, de por sí, un inconveniente, pero el problema se agrava aún más cuando pretendemos que el modelo resulte válido hasta 10^{-45} segundos después de la gran explosión. Los agregados incipientes de materia se desarrollan rápidamente con el tiempo en virtud de su propia atracción gravitacional; por consiguiente, un modelo que se aplique a tiempos muy tempranos empezará con inhomogeneidades muy pequeñas. Para comenzar a los 10^{-45} segundos, la materia ha de partir de un estado peculiar con una extraordinaria uniformidad, aunque no absoluta. Un gas normal, en equilibrio térmico, resultaría poco homogéneo,

debido a los movimientos aleatorios de sus partículas. Esta peculiaridad del estado inicial de la materia, exigida por el modelo normal, recibe el nombre de problema de la suavidad.

La densidad energética del universo plantea otra sutil dificultad al modelo normal. De acuerdo con la relatividad general, el espacio del universo puede ser curvo; el tipo de curvatura dependerá de la densidad energética. Si ésta supera cierto valor crítico, que depende de la velocidad de expansión, diremos que el universo es cerrado: el espacio se curva sobre sí mismo y forma un volumen finito sin ninguna frontera. (Un símil familiar es la superficie de una esfera, que tiene un área finita, sin frontera.) Con densidad energética menor que la densidad crítica, nos hallaremos ante un universo abierto: el espacio se curva, aunque no sobre sí mismo, y el volumen es infinito. Si la densidad energética es exactamente igual a la densidad crítica, el universo será plano: la habitual geometría de Euclides describiría este espacio (de nuevo el volumen es infinito).

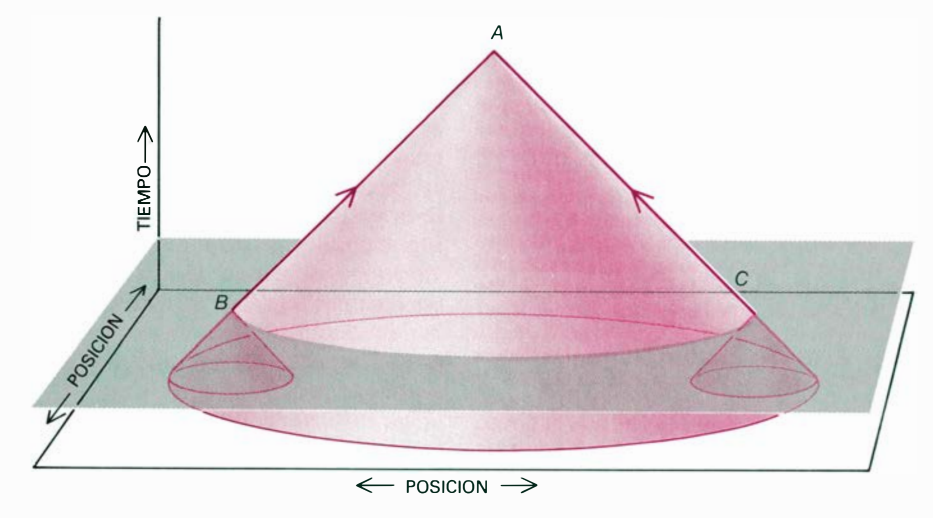
El cociente entre la densidad energética del universo y la densidad crítica es una cantidad que los cosmólogos designan con la letra griega Ω (omega). El valor $\Omega = 1$ (que corresponde a un universo plano) representa un estado de equilibrio inestable. Si Ω llegó en algún instante a valer exactamente la unidad, habría permanecido siempre exactamente igual a uno. Sin embargo, si Ω difirió ligeramente de la unidad un instante después de la explosión inicial, esta desviación respecto de uno creció rápidamente con el tiempo. Ante semejante inestabilidad, sorprende que el valor actualmente medido para Ω esté entre 0,1 y 2,0. (Los cosmólogos no están todavía seguros de si el universo es abierto, cerrado o plano.) Para que el valor de Ω esté hoy incluido en este estrecho margen, su valor un segundo después de la explosión inicial tenía que haber diferido de la unidad en una parte sobre 10^{15}. El modelo normal no ofrece ninguna razón para que Ω empezara con un valor tan cercano a uno; simplemente toma esta hipótesis como condición inicial. Este defecto del modelo normal, llamado problema de la curvatura nula, fue señalado, en primer lugar, en 1979, por Robert H. Dicke y P. James E. Peebles, de Princeton.

Las ventajas e inconvenientes del modelo de la gran explosión que hemos considerado hasta ahora afectan a la cosmología, la astrofísica y la física nuclear. Ahora bien, si intentamos aplicar el modelo de la gran explosión a tiempos más remotos, nos moveremos en una época en que esas partes de la física se ven rebasadas. En ese intervalo temporal, toda la materia estaría descompuesta en las partículas elementales constituyentes. Al objeto de abordar esa época, los cosmólogos han recurrido a los últimos resultados obtenidos en la teoría de partículas elementales. No hay duda de que uno de los principales avances de los últimos diez años ha sido la unión de intereses entre la física de partículas, la astrofísica y la cosmología. Las consecuencias para el modelo de la gran explosión conducen, como mínimo, a un nuevo éxito y, como mínimo también, a un fracaso.

El desarrollo tal vez más importante de la teoría de partículas elementales corresponde, en estos últimos diez años, a la noción de teorías de gran unificación, cuyo prototipo fue propuesto en 1974 por Howard M. Georgi y Sheldon Lee Glashow, de la Universidad de Harvard. Se trata de teorías de difícil verificación experimental, ya que sus predicciones más significativas se refieren a energías muy superiores a las que se pueden alcanzar en nuestros aceleradores de partículas. Con todo, estas teorías tienen algún pilar experimental y unifican tan bien la comprensión de las interacción de las partículas elementales que muchos físicos las consideran del máximo interés.

La idea fundamental que guía a las teorías de gran unificación es que la unicidad en una sola fuerza real de las que se consideraban tres fuerzas independientes: la fuerte, la débil y la electromagnética. La simetría relaciona las fuerzas entre sí. Puesto que la experiencia nos muestra a estas fuerzas muy distintas por su intensidad y su carácter, la teoría unificada se construirá de suerte que la simetría se rompa espontáneamente al aplicarla al universo actual.

Una simetría rota espontáneamente es aquella que está presente en la teoría fundamental que describe un sistema, aunque subyace escondida en el estado de equilibrio del sistema. Por ejemplo, un líquido descrito por leyes físicas que tienen simetría rotacional también tiene simetría rotacional: la distribución de moléculas nos aparece idéntica aunque giremos el líquido. Sin embargo, cuando el líquido se solidifica dando un cristal, los átomos se ordenan según los ejes cristalográficos y la simetría se rompe. Cabe, pues, esperar que, si elevamos la temperatura de un estado de un sistema con la simetría rota, podamos lograr una transición de fase hacia un estado que recupere la simetría, del mismo modo que un cristal se funde y forma un líquido. Las teorías de gran unificación predicen esta transición de fase a una temperatura crítica de unos 10^{27} grados.



3. EL PROBLEMA DEL HORIZONTE es un inconveniente serio para la teoría normal o estándar de la gran explosión. En este diagrama espacio-tiempo tridimensional, las escalas se han tomado de modo no lineal; así pues, la trayectoria de un pulso de luz se corresponde con una línea que forma un ángulo de 45 grados con el eje vertical. El punto *A* indica nuestra posición en el espacio y en el tiempo. Como ninguna señal puede propagarse a velocidades superiores a la de la luz, sólo recibiremos señales de la región coloreada: nuestro cono de luz hacia el pasado. Los sucesos que caen fuera del cono de luz del pasado de un punto dado no pueden afectar al suceso que ocurra en dicho punto. El plano horizontal gris indica el tiempo en que se liberó la radiación de fondo de microondas. La radiación, que ahora nos llega de direcciones opuestas, fue liberada en los puntos *B* y *C*; desde allí, siguiendo nuestro cono del pasado, ha llegado hasta el punto *A*. El cono del pasado del punto *B* no tiene ninguna intersección con el cono del pasado del punto *C* y, por tanto, los dos puntos no han estado sometidos a influencias mutuas. El problema del horizonte reside en la dificultad de explicar cómo la radiación recibida desde dos sentidos opuestos llegan con la misma temperatura. En el modelo normal, la uniformidad de la temperatura, evidenciada en la radiación de fondo de microondas, debe imponerse como condición inicial del universo.

Una propiedad nueva de las teorías de gran unificación concierne a las partículas llamadas bariones, clase cuyos miembros más importantes son el protón y el neutrón. Hasta ahora, en todos los procesos físicos observados, el número de bariones menos el número de antibariones no cambia; en el lenguaje de la física de partículas decimos que se conserva el número bariónico total del sistema. En virtud de esta ley de conservación, el protón debe ser totalmente estable; por tratarse del barión más ligero, no puede desintegrarse en otras partículas sin cambiar el número bariónico total. Merced a la experimentación, sabemos que la vida media del protón trasciende los 10^{31} años.

TIPO DE UNIVERSO	COCIENTE ENTRE LA DENSIDAD DE ENERGIA Y LA DENSIDAD CRITICA (Ω)	GEOMETRIA ESPACIAL	VOLUMEN	EVOLUCION TEMPORAL
CERRADO	>1	CURVATURA POSITIVA (ESFERICA)	FINITO	SE EXPANDE Y RECOLAPSA
ABIERTO	<1	CURVATURA NEGATIVA (HIPERBOLICA)	INFINITO	SE EXPANDE SIEMPRE
PLANO	1	CURVATURA NULA (EUCLIDEA)	INFINITO	SE EXPANDE SIEMPRE PERO LA VELOCIDAD DE EXPANSION TIENDE A CERO

4. TRES TIPOS DE UNIVERSO: cerrado, abierto y plano. Pueden deducirse del modelo normal, o estándar, de la gran explosión (incluida la hipótesis usual de la inalteración de las ecuaciones de la relatividad general al agregar un término cosmológico). Lo que distingue las diferentes geometrías es el valor de la cantidad llamada Ω, que es el cociente entre la densidad de energía del universo y una densidad llamada crítica, cuyo valor depende, a su vez, de la intensidad de la expansión del universo. El valor actualmente admitido para Ω está entre 0,1 y 2,0; ello implica que su valor, un segundo después de la gran explosión, era igual a 1, con un error de una parte entre 10^{15}. La dificultad que encierra explicar este valor de Ω en el modelo de la gran explosión recibe el nombre de problema de la curvatura plana.

Las teorías de gran unificación implican que el número bariónico no se conserve con total exactitud. Cuando la temperatura es baja, en la fase de simetría rota, la ley de conservación constituye un enfoque excelente; el límite observado de la vida media del protón resulta entonces compatible con muchas versiones de las teorías de gran unificación. Sin embargo, a temperaturas elevadas se espera que sean frecuentes los procesos en los que cambie el número bariónico de un sistema de partículas.

De la fusión del modelo de la gran explosión con las teorías de gran unificación se infiere, directamente, la asimetría entre materia y antimateria que hay en el universo. Se cree que cuantas estrellas, galaxias y polvo se han observado en el universo lo están más en forma de materia que de antimateria; sus partículas nucleares incluyen más bariones que antibariones. De ello se deduce que el número bariónico total del universo es del orden de 10^{78}. Antes de que aparecieran las teorías de gran unificación, en que se admitía que el número bariónico debía conservarse, había que postular también este total bariónico como otra de las condiciones iniciales del universo. Mas, cuando combinamos las teorías de gran unificación con la descripción de la gran explosión, el exceso observado de la materia sobre la antimateria puede producirse naturalmente por medio de las interacciones entre partículas elementales a temperaturas un poco inferiores a la temperatura crítica de la transición de fase. En las teorías de gran unificación, los cálculos dependen de muchísimos parámetros arbitrarios que impiden avanzar una predicción cuantitativa; a pesar de ello, la asimetría materia-antimateria observada puede producirse mediante una elección razonable de valores para estos parámetros.

De la combinación de las teorías de gran unificación con la descripción de la gran explosión resulta un problema serio: el gran número de defectos que suelen aparecer durante la transición de la fase simétrica a la fase de simetría rota. Estos defectos surgen cuando las regiones que están en la fase simétrica sufren una transición a estados diferentes de simetría rota. Si, en una situación análoga, cristaliza un líquido, otras regiones diferentes pueden iniciar su cristalización con orientaciones diferentes de sus ejes cristalográficos. Los dominios con distinta orientación cristalográfica crecen y se unen; para ellos, resulta favorable, desde el punto de vista energético, la suavización de las diferencias de orientación en sus fronteras. Se trata, empero, de una suavización imperfecta, por lo común; y quedan defectos locales.

En las teorías de gran unificación, los defectos puntuales y los superficiales plantean serios problemas cosmológicos. Corresponden los primeros a monopolos magnéticos; los defectos superficiales son las llamadas paredes del dominio. Ambos deben ser muy estables y de gran masa. (Puede demostrarse que el monopolo pesa unas 10^{16} veces más que el protón.) Un dominio intercorrelacionado en la fase de simetría rota no puede ser mucho mayor que la distancia del horizonte correspondiente a este tiempo y, de ahí, puede calcularse el número de defectos que se producen durante la transición. De lo que resulta que habría tantos defectos después de la transición que su masa dominaría la densidad de energía del universo y, por tanto, debería acelerar la evolución subsiguiente. La radiación de fondo de microondas habría alcanzado la temperatura actual de tres grados K transcurridos sólo 30.000 años de la gran explosión, y no a los 10.000 millones de años; se esfumarían así todas las predicciones acertadas del modelo de la gran explosión. Por tanto, toda unión próspera de las teorías de gran unificación con la descripción de la gran explosión tendrá que incorporar algún mecanismo que suprima drásticamente la producción de monopolos magnéticos y paredes de dominio.

El modelo del universo inflacionario ofrece una solución satisfactoria a estos problemas. Pero antes de explicarlo, nos detendremos brevemente en la ruptura de la simetría y en las transiciones de fase en las teorías de gran unificación.

Todas las teorías modernas sobre las partículas, incluidas las teorías de gran unificación, pertenecen a la clase de teorías cuánticas de campo. De éstas, la más conocida es la que describe el electromagnetismo. Según la teoría clásica (no cuántica) del electromagnetismo, desarrollada por James Clerk Maxwell, hacia 1860, los campos eléctrico y magnético tienen un valor perfectamente definido en cada punto del espacio; un conjunto dado de ecuaciones describe su variación con el tiempo. En los primeros años del siglo XX, la teoría de Maxwell se retocó para hacerla compatible con la teoría cuántica. En la teoría clásica, se puede añadir cualquier cantidad de energía al campo electromagnético; en la teoría cuántica, por contra, la energía sólo puede aumentar en cantidades discretas, los cuantos, que, en este caso, reciben el nombre de fotones. Los fotones tienen propiedades de onda de partícula; en el léxico de la física moderna se les designa habitualmente como partículas. En general, una

teoría de campos cuántica se basa en una teoría de campos clásica que se transforma en una teoría de partículas cuando aplicamos las reglas de la teoría cuántica.

Ingrediente esencial de las teorías de gran unificación, lo indicamos antes, es el fenómeno de la ruptura espontánea de la simetría. El mecanismo a través del cual se produce la ruptura espontánea de la simetría en las teorías de gran unificación resulta, en muchos aspectos, más sencillo que en el caso de los cristales. Se produce la ruptura espontánea de la simetría en una teoría de gran unificación cuando en la formulación de la teoría introducimos un conjunto especial de campos, conocidos por campos de Higgs (en honor de Peter W. Higgs, de la Universidad de Edimburgo). No se quiebra la simetría cuando todos los campos de Higgs son nulos, pero se rompe espontáneamente cuando uno al menos de dichos campos toma valores no nulos. Además, podemos formular la teoría de tal modo que un campo de Higgs tenga un valor no nulo en el estado de densidad de energía más bajo, que, en este contexto, recibe el nombre de vacío verdadero. Para temperaturas superiores a los 10^{27} grados, las fluctuaciones térmicas dan el cero como valor de equilibrio del campo de Higgs; con ello se asiste a una transición hacia la fase simétrica.

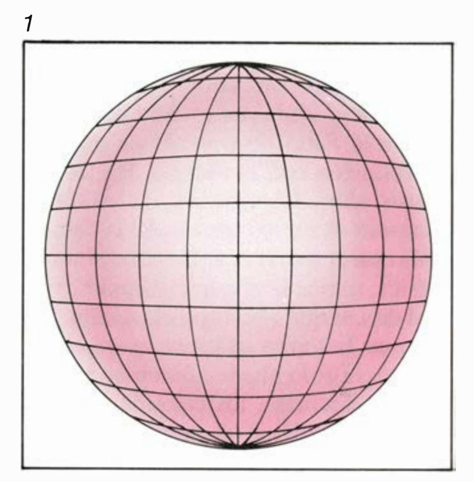
Ya hemos reunido información suficiente para acometer el examen del modelo inflacionario del universo. Empezaremos por la forma en que fue propuesto por uno de nosotros (Guth) en 1980. Todo modelo cosmológico debe arrancar con algunas hipótesis sobre las condiciones iniciales; mas, para el modelo inflacionario, las condiciones iniciales pueden ser bastante arbitrarias. Sin embargo, debemos suponer que el universo primitivo contenía, al menos, algunas regiones de gas muy caliente comparado con la temperatura crítica de la transición de fase y, además, en expansión. En tales regiones calientes, el campo de Higgs sería nulo. En virtud del descenso de la temperatura provocado por la propia expansión, se producirían circunstancias termodinámicas favorables para que el campo de Higgs tomara valores no nulos; ello llevaría el sistema a su fase de simetría rota.

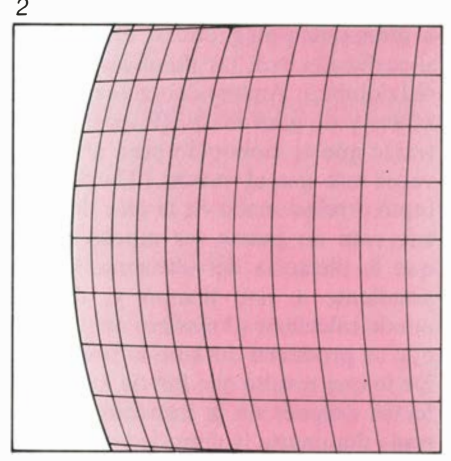
Para algunos valores de los parámetros desconocidos de las teorías de gran unificación, la transición de fase transcurriría muy lentamente, si la comparamos con la velocidad de enfriamiento. Lo que implicaría que el sistema podría enfriarse bastante por debajo de los 10^{27} grados, con el campo de Higgs todavía nulo. Este fenómeno, de sobreenfriamiento, es muy corriente en la física de la materia condensada. (Recuérdese que el agua puede sobreenfriarse hasta 20 grados por debajo del punto de congelación; otro ejemplo: el vidrio se consigue mediante un rápido sobreenfriamiento de un líquido hasta una temperatura claramente inferior a su punto de congelación.)

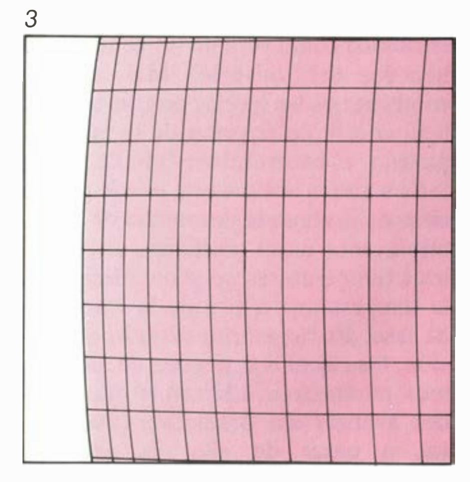
Si la región que contiene el gas continuase enfriándose, tendería hacia un estado peculiar de la materia, llamado falso vacío. Aunque este estado de la materia no se ha observado nunca, tiene propiedades que se deducen nítidamente de la teoría cuántica de campos. La temperatura y, por tanto, la componente térmica de la densidad de energía descenderían rápidamente; la densidad de enegía del estado se concentraría totalmente en el campo de Higgs. Un valor nulo del campo de Higgs implica una gran densidad de energía para el falso vacío. En la forma clásica de la teoría, dicho estado sería absolutamente estable, aun cuando no constituyera el estado de mínima densidad de energía. Los estados con densidad de energía más baja se verían separados del falso vacío por una barrera intermedia de energía y no habría energía disponible para permitir que el campo de Higgs superara la barrera [*véase la figura 2, arriba*].

En la versión cuántica del modelo, el falso vacío no es absolutamente estable. Según las reglas de la teoría cuántica, todos los campos se hallan en continua fluctuación. Como indicara ya Sidney R. Coleman, de Harvard, una fluctuación cuántica podría ser la responsable ocasional de que el campo de Higgs de una pequeña región del espacio atravesase por "efecto túnel", la barrera de energía, sirviendo de núcleo para una "burbuja" de la fase de simetría rota. A partir de aquí, la burbuja empezaría a crecer a una velocidad que, rápidamente, se acercaría a la velocidad de la luz, transformando el falso vacío en la fase de simetría rota. La intensidad con la que se forman burbujas depende, en gran medida, del valor de los parámetros desconocidos de la teoría de gran unificación; en el modelo inflacionario se supone que esta intensidad es muy baja.

La propiedad más característica del falso vacío quizá sea el valor de la presión, grande y negativo a la vez. Para conocer su razón, volvamos al proceso por el que se forma una burbuja de verdadero vacío, para luego crecer dentro de una región del falso vacío. Este crecimiento viene energéticamente favorecido, ya que el vacío verdadero tiene







5. SOLUCION AL PROBLEMA DE LA CURVATURA PLANA. Dedúcese de esta serie de dibujos en perspectiva de una esfera que está hinchándose. La composición gráfica ilustra cómo puede alcanzarse una geometría espacial plana (que corresponde al valor 1 de Ω, símbolo que se explica en la figura 4), a partir del modelo inflacionario, de un modo sencillo y natural. De uno a otro dibujo, de ambas páginas, la esfera se hincha por un factor tres (a su vez, el

una densidad de energía más baja que el falso. Sin embargo, el crecimiento nos indica también que la presión del verdadero debe ser mayor que la del falso vacío, obligando a la burbuja a aumentar de tamaño. Por ser nula la presión del vacío verdadero, la presión del vacío falso tendrá que ser negativa. Una argumentación más detallada muestra que la presión del vacío falso es igual al valor negativo de su densidad de energía (cuando medimos con las mismas unidades las dos cantidades).

Esta presión negativa no provoca la aparición de fuerzas mecánicas dentro del falso vacío, ya que las fuerzas mecánicas sólo aparecen cuando hay diferencias de presión. Sí habría efectos gravitacionales. En circunstancias ordinarias, la expansión de la región que contiene el gas sería frenada por la mutua atracción gravitacional de la materia que está dentro de ella. En la física de Newton, esta atracción es proporcional a la densidad de masa, que, en las teorías relativistas, es igual a la densidad de energía dividida por el cuadrado de la velocidad de la luz. Según la relatividad general, la presión contribuye también a la atracción; más en concreto, la fuerza gravitacional es proporcional a la densidad de energía más tres veces la presión. Para el falso vacío, la contribución de la presión sobrepasaría con mucho la contribución de la densidad de energía y tendría signo opuesto. De aquí que, de la noción extraña de presión negativa, lleguemos al efecto todavía más extraño de una fuerza gravitacional que es en realidad repulsiva. En consecuencia, la expansión de la región quedaría acelerada y crecería exponencialmente, doblando su diámetro cada intervalo de unos 10^{-34} segundos.

Este período de expansión acelerada recibe el nombre de era inflacionaria; es el elemento clave del modelo inflacionario del universo. Según éste, la era inflacionaria duró unos 10^{-32} segundos o más, a lo largo de cuyo período de diámetro del universo se multiplicó por un factor 10^{50} o superior. Se presupone que, tras esta colosal expansión, se produjo la transición a la fase de simetría rota. Se liberó, entonces, la densidad de energía del falso vacío, dando lugar a una enorme producción de partículas. La región se recalentó hasta una temperatura de unos 10^{27} grados. (En termodinámica, a la energía liberada se le da el nombre de calor latente; es análoga a la energía liberada cuando se congela el agua.) A partir de este punto, la región continuaría expandiéndose y enfriándose a la velocidad correspondiente a la del modelo normal de la gran explosión. Un volumen igual al del universo observable cabría bien dentro de una tal región.

El problema del horizonte queda eliminado de una manera sencilla. En el modelo inflacionario, el universo observado procede de una región cuyo diámetro es mucho menor (por un factor de 10^{50} o superior) que el de la correspondiente región del modelo normal. Antes de que empiece la inflación, la región es mucho menor que la distancia del horizonte y tiene tiempo para homogeneizarse y alcanzar el equilibrio térmico. Esta pequeña región homogénea sufre luego la inflación y se agranda hasta abarcar el universo observado. Por tanto, las fuentes de la radiación de fondo de microondas, que ahora nos llega de todas las direcciones, estuvieron antes muy próximas unas de otras; dispusieron del tiempo necesario para alcanzar una temperatura común antes de que empezara la era inflacionaria.

El problema de la curvatura nula queda también solventado de un modo sencillo y natural. Las ecuaciones que describen la evolución del universo durante la era inflacionaria son distintas de las del modelo normal. Sucede que el valor del cociente Ω tiende rápidamente a uno, sea cual sea el valor que poseyera antes de la inflación. Comportamiento que entenderemos fácilmente si recordamos que el valor de $\Omega = 1$ corresponde a un espacio geométricamente plano. La rapidez de la expansión provoca el aplanamiento del espacio, igual que la superficie de un globo se aplana cuando se hincha. El mecanismo responsable de que Ω tienda a uno es tan eficaz que conduce a una predicción casi exacta: el valor actual de Ω debe ser prácticamente igual a uno. Aunque muchos astrónomos (no todos, sin embargo) creen que el valor uno es compatible con nuestras observaciones, una determinación más precisa de Ω conduciría a una prueba crucial para el modelo inflacionario.

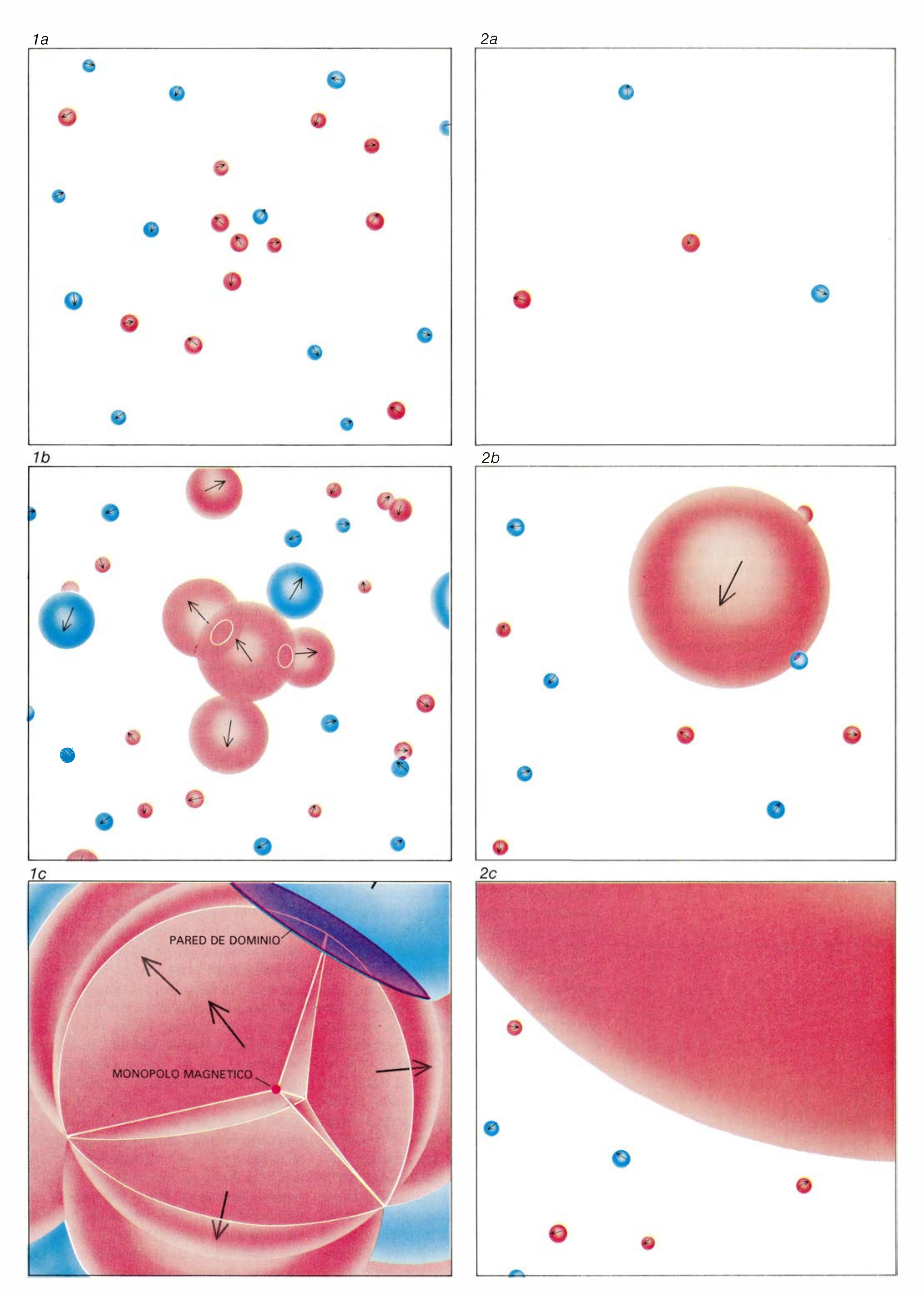
En su formulación original, el modelo inflacionario tenía un defecto grave: en las condiciones descritas, la propia transición de fase crearía inhomogeneidades mucho más acusadas que las que actualmente se observan. Como ya se ha explicado, la transición de fase ocurre con la creación aleatoria de núcleos de burbujas de la nueva fase. Puede demostrarse que las burbujas permanecen siempre agrupadas en cúmulos finitos, desconectados unos de otros; en cada cúmulo habría una burbuja mayor que las demás y sería la dominante. Inicialmente, casi toda la energía del cúmulo estaría concentrada en la superficie de la gran burbuja; no existe, que se sepa, ningún mecanismo que redistribuya uniformemente la energía. Esta configuración no presenta ninguna semejanza con el universo observado.

Durante los dos años subsiguientes a la aparición del modelo, éste se erigió en una solución tentadora, aunque imperfecta a la hora de hacer frente a distintos problemas cosmológicos. Sin embargo, a finales de 1981, surgió un nuevo intento, empeñado por A. D. Linde, del Instituto de Física P.N. Lebedev de Moscú, y, de modo independiente, por Andreas Albrecht y uno de nosotros (Steinhardt), de la Universidad de Pennsylvania. El planteamiento en cuestión, conocido por nuevo universo inflacionario, mantiene todas las ventajas y evita todos los inconvenientes del modelo original.

Para el nuevo planteamiento, lo decisivo es dar una forma especial a la función densidad de energía que describe el campo de Higgs [*véase la figura 2, centro*]. De las teorías de campo cuánticas con funciones de densidad de energía de este tipo se había ocupado ya Coleman, en colaboración con Erick J. Weinberg, de la Universidad de Columbia. A diferencia de lo que implica el caso más típico, recogido en la ilustración superior de la figura 2, no hay ninguna barrera de energía que separe el falso vacío del verdadero; el falso vacío se halla ahora en la cima de una meseta casi llana. En el contexto de las teorías de gran unificación, se requiere una elección concreta de los valores de los parámetros para obtener esta función de densidad de energía. Función

4

número de líneas grises en la superficie lo aumentamos por ese factor). La curvatura de la superficie se hace despreciable en la escala del dibujo.


1a
2a
1b
2b
1c
2c
PARED DE DOMINIO
MONOPOLO MAGNETICO

de densidad de energía que, lo explicaremos enseguida, da un comportamiento especial a la transición de fase; se habla entonces de transición de caída lenta en rotación.

La situación al principio es la misma que la del modelo inflacionario original. Como antes, debemos suponer que en el universo primitivo había regiones calientes con temperaturas superiores a los 10^{27} grados y, además, en expansión. En tales regiones, las fluctuaciones térmicas harán que el valor de equilibrio de los campos de Higgs sea cero y no habrá ruptura de la simetría. A medida que la temperatura desciende, resultará termodinámicamente favorable para el sistema sufrir una transformación de fase en la que por lo menos uno de los campos de Higgs adquiera un valor no nulo, pasando a la fase de simetría rota. Sin embargo, como en el caso anterior, la intensidad de esta transformación de fase será muy baja, comparada con la intensidad de enfriamiento. El sistema se sobreenfriará hasta una temperatura despreciable, manteniendo el campo de Higgs su valor cero; el estado resultante constituirá, de nuevo, un falso vacío.

En el modo en que tiene lugar la transición de fase hallamos una peculiaridad notable del nuevo modelo. Las fluctuaciones cuánticas o pequeñas fluctuaciones térmicas residuales provocarían que el campo de Higgs dejase de ser nulo. Al no haber barrera de energía, el valor del campo de Higgs irá creciendo uniformemente; la intensidad de este crecimiento se asimila a la que presenta la velocidad de una bola que rodara por una colina del mismo perfil que la curva que representa la densidad de energía, sometida a una fuerza de rozamiento. Como dicha curva es casi plana en los alrededores del punto en el que el campo de Higgs se anula, la evolución en sus primeras etapas es muy lenta. Mientras el campo de Higgs esté cerca del valor cero, la densidad de energía será prácticamente igual a la del falso vacío. Como en la situación original, la región sufrirá una expansión acelerada, doblando sucesivamente su diámetro cada 10^{-34} segundos. Sin embargo, ahora, la expansión dejará de acelerarse cuando el valor del campo de Higgs llegue a la parte más pronunciada de la curva. Podemos determinar cuál será la inflación calculando el tiempo que precisa el campo de Higgs para hacer el recorrido. Creemos muy plausible un factor de expansión del orden de 10^{50} o superior, si bien el valor concreto dependerá de los pormenores de la teoría de partículas que se escoja.

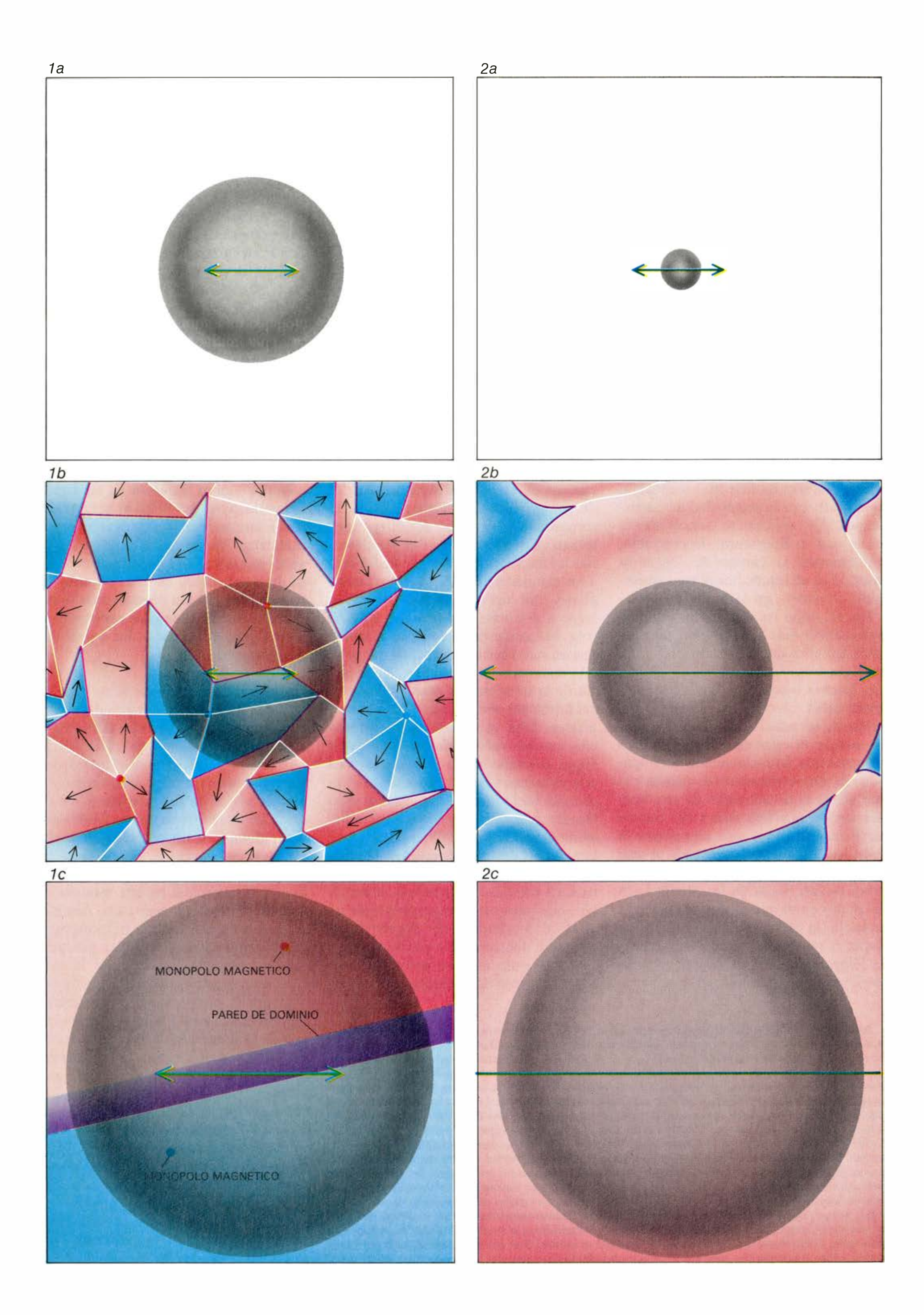
Hasta aquí, hemos simplificado un tanto la descripción de la transición de fase. La verdad es que hay muchos estados de simetría rota, como posibles orientaciones hay para los ejes de un cristal. Hay, asimismo, bastantes campos de Higgs; los distintos estados de simetría rota se distinguen por el conjunto de campos de Higgs que toman valores no nulos. Por ser aleatorias las fluctuaciones que sacan del cero a los campos de Higgs, las distintas regiones del universo primitivo darán lugar a diferentes estados de simetría rota, formando cada región un dominio con un radio inicial prácticamente igual a la distancia del horizonte. Al principio de la transición de fase, la distancia del horizonte será de unos 10^{-24} centímetros. Una vez constituido el dominio, con los valores de los campos de Higgs desviándose un poco de cero en una determinada combinación, evolucionará hacia uno de los estados estables de simetría rota y sufrirá una inflación por un factor de 10^{50} o superior. El tamaño del dominio después de la inflación superará, pues, los 10^{26} centímetros. La totalidad del universo observable, que para ese tiempo sería sólo de unos 10 centímetros de diámetro, cabría bien dentro de un solo dominio.

Durante esa enorme inflación, cualquier densidad de partículas que hubiera en un comienzo quedará reducida prácticamente a cero. La energía contenida en la región se debería en su totalidad a la energía almacenada en el campo de Higgs. ¿Cómo puede liberarse esta energía? Una vez el campo de Higgs se ha alejado de la zona llana de la curva de densidad de energía, empezará a oscilar rápidamente alrededor del valor del vacío verdadero. Basándonos en la relación entre partículas y campos de la teoría cuántica de campos, esta situación puede describirse también como un estado con gran densidad de partículas de Higgs. Ahora bien, las partículas de Higgs serían inestables: se desintegrarían rápidamente cuando otras más ligeras, que a su vez interaccionarían entre sí y posiblemente sufrirían nuevas desintegraciones. El sistema se convertiría muy pronto en un gas caliente de partículas elementales en equilibrio térmico, tal como habíamos supuesto en las condiciones iniciales del modelo normal. La temperatura de recalentamiento puede calcularse, dando un valor entre 2 y 10 veces inferior al de la temperatura crítica de la transición de fase. A partir de aquí, la descripción coincide con la del modelo normal de la gran explosión; se mantienen, pues, todos los logros del modelo normal.

Adviértase que el punto débil del modelo inflacionario original se evita cuidadosamente. Simplificando, podemos decir que las burbujas aisladas que aparecían en el modelo original quedan reemplazadas ahora por los dominios. Los dominios de la lenta transición de caída en rotación estarían rodeados por otros dominios, no por el falso vacío; no tienden hacia la esfericidad. Se evita, pues, el término "burbuja". La diferencia principal en este modelo inflacionario está en que cada dominio sufre la inflación durante su formación, dando lugar a una gran región prácticamente homogénea dentro de la cual cabe el universo observable.

Como la temperatura de recalentamiento es cercana a la temperatura crítica de la transición de fase de la teoría de gran unificación, la asimetría materia-antimateria se produciría por interacciones entre partículas, inmediatamente después de la transición de fase. El procedimiento de producción es el mismo que el que se deduce de las teorías de gran unificación para el modelo normal de la gran explosión. Sin embargo, a diferencia de lo que pasa en este modelo estándar, en el modelo inflacionario no cabe la posibilidad de imponer, como condición inicial, el nú-

6. BURBUJAS EN EXPANSION, pertenecientes a la fase de simetría rota; se forman en una región en expansión de la fase simétrica y están aquí dibujadas en dos secuencias temporales muy esquemáticas. La secuencia correspondiente al modelo normal de la gran explosión (*a la izquierda*) abarca un tiempo mucho más corto que el de la secuencia que representa la forma original del modelo inflacionario (*derecha*). En ambos casos, los campos de Higgs tienen valor nulo en la región exterior a las burbujas, mientras que, al menos, un campo de Higgs tiene un valor no nulo dentro de cada burbuja. Los estados de simetría rota en una teoría de gran unificación se caracterizan mediante parámetros de dos tipos: discreto y continuo. En la ilustración, cada burbuja porta doble contraseña: el color (azul o rojo), para señalar el parámetro discreto y una flecha negra en el interior para indicar el valor del parámetro continuo. En el modelo normal, las burbujas se unen y completan la transición de la fase simétrica a la fase de simetría rota. Un defecto superficial, llamado pared de dominio, se puede formar en todas las fronteras entre regiones con valores diferentes del parámetro discreto (*zonas en púrpura*). En las regiones con color uniforme, surgen defectos puntuales, llamados monopolos magnéticos, creados en centros de intersección de varias burbujas, siempre que la flecha que representa el parámetro continuo apunte hacia hacia cualquier dirección que no sea el centro. En la forma original del modelo inflacionario, la expansión rápida de la región de falso vacío, o fase simétrica, evitaría que las burbujas se unieran. Las dos situaciones hipotéticas tienen consecuencias contrarias a la observación; el modelo inflacionario apareció para evitar estas dificultades.


1a
2a
1b
2b
1c
2c
MONOPOLO MAGNETICO
PARED DE DOMINIO
MONOPOLO MAGNETICO

mero bariónico total del universo; la inflación que luego ocurre diluiría cualquier densidad inicial del número bariónico hasta valores inapreciables. Por consiguiente, la viabilidad del modelo inflacionario depende, de una manera decisiva, de la viabilidad de las teorías de partículas que, como las teorías de gran unificación, no conservan el número bariónico.

Ya nos encontramos en disposición de buscar soluciones a los problemas cosmológicos que antes hemos planteado. Los problemas del horizonte y de la curvatura plana se resuelven por el mismo mecanismo que aparece en el modelo original de universo inflacionario. Con la nueva teoría inflacionaria podemos resolver también el problema relativo a los monopolos y a las paredes del dominio. Estos defectos se formarían en las fronteras que separan los dominios; ahora bien, estos últimos adquirían por inflación tamaños enormes haciendo que los defectos cayeran a distancias muy por encima del alcance de nuestras observaciones. (Por efecto térmico posterior a la transición se generarían algunos defectos más, pero se cree que su número sería despreciable.)

De este modo, con unas pocas ideas sencillas, el modelo mejorado del universo inflacionario aportaría una solución correcta a problemas difíciles que se presentan en la descripción normal de la gran explosión: los relativos al horizonte, la curvatura plana, el monopolo magnético y las paredes de dominio. Por desgracia, la necesaria transición de caída lenta en rotación exige un ajuste preciso de los parámetros; los cálculos dan predicciones aceptables sólo si los parámetros toman valores dentro de un estrecho margen. A la mayoría de los físicos teóricos, entre los que nos incluimos, les parece improbable un ajuste tan preciso. Sin embargo, las consecuencias de la descripción son tan acertadas que nos sentimos animados a seguir, con la esperanza de encontrar versiones realistas de las teorías de gran unificación en las que una transición de caída lenta en rotación no requiera un ajuste fino.

Los éxitos reseñados dan un aspecto convincente al nuevo modelo inflacionario. Se acaba de descubrir, además, que el modelo puede resolver también un problema cosmológico adicional que no se había aún considerado cuando se desarrolló el modelo: el problema de la suavidad. La aparición de faltas de homogeneidad en la densidad dentro del nuevo universo inflacionario fue presentada, durante el verano de 1982, en el Nuffield Workshop sobre el universo muy primitivo; de ellas se ocuparon, entre otros, James M. Bardeen, de la Universidad de Washington, Stephen W. Hawking, de Cambridge, So-Young Pi, de Boston, Michael S. Turner, de Chicago, A. A. Starobinsky, del Instituto de Física Teórica L.D. Landau de Moscú, y los autores. Se vio que el nuevo modelo inflacionario, a diferencia de los anteriores, auspiciaba una predicción precisa del espectro de las inhomogeneidades. A grandes rasgos, el proceso de inflación suaviza primero cualquier inhomogeneidad primordial que pudiera existir en las condiciones iniciales. Luego, durante la transición de fase, se producen fallos en la homogeneidad instados por las fluctuaciones cuánticas del campo de Higgs de un modo tal que queda fijado completamente por las leyes físicas subyacentes. Las inhomogeneidades surgen a una escala de longitud muy pequeña, en la que son importantes los fenómenos cuánticos; luego, a través del proceso inflacionario, crecen hasta escalas astronómicas.

La forma prevista para el espectro de inhomogeneidades destaca por la invariancia de escala; es decir, la magnitud de las inhomogeneidades es casi igual para todas las escalas de longitud con significado astrofísico. Esta predicción apenas si depende de los pormenores de la teoría de gran unificación en que se base. Un espectro de esta misma forma fue ya propuesto a principios de los años setenta, como modelo fenomenológico para la creación de galaxias, por Edward R. Harrison, de la Universidad de Massachusetts en Amherst, y Yakov B. Zel'dovich, del Instituto de Problemas Físicos de Moscú, en trabajos independientes. Los detalles sobre la formación de galaxias son complejos y todavía no bien comprendidos, pero muchos cosmólogos creen que es necesario un espectro de inhomogeneidades con invariancia de escala para explicar cómo se alcanzó la estructura actual de las galaxias y de los cúmulos de galaxias [véase "Macroestructura del universo", por Joseph Silk, Alexander S. Szalay y Yakov B. Zel'dovich; INVESTIGACIÓN Y CIENCIA, diciembre de 1983].

El nuevo modelo inflacionario predice también la magnitud de las inhomogeneidades de la densidad. Se trata, empero, de una predicción muy sensible a la exposición de la teoría de partículas que se utilice. Desgraciadamente, el valor que se obtiene a partir de la teoría de gran unificación más sencilla es demasiado grande para corresponderse con la uniformidad que se observa en la radiación cósmica de fondo de microondas. No obstante el problema que representa esta falta de coherencia, seguimos sin saber todavía si la teoría de gran unificación más sencilla es la correcta. En concreto, la teoría de gran unificación más sencilla predice una vida media para el protón menor que el actual límite experimental. Por otro lado, pueden construirse teorías de gran unificación más complicadas que den inhomogeneidades de la densidad del valor deseado. Muchos investigadores creen que cuando se tenga una teoría de partículas correcta, el nuevo modelo inflacionario añadirá a sus éxitos la solución del problema de la suavidad.

Las teorías de campo cuánticas con una nueva simetría llamada supersimetría presentan una línea de investigación prometedora. La supersimetría relaciona las propiedades de las partículas con momento angular entero con las de las partículas con momento angular semientero; de ahí que limite fuertemente la forma de la teoría. Muchos físicos teóricos creen necesaria la supersimetría para construir una teoría consistente de la gravedad cuántica y para, finalmente, unificar la gravedad con las fuerzas fuerte, débil y electromagnética. Una tentadora propiedad de los modelos que admiten la supersimetría

7. EL NUEVO MODELO INFLACIONARIO elude hábilmente los problemas del horizonte, de los monopolos magnéticos y de las paredes de dominio. En las dos series de dibujos de la página contigua, correspondientes al modelo normal de la gran explosión (*izquierda*) y al nuevo modelo inflacionario (*derecha*), la esfera gris indica la región del espacio cuya evolución dará el universo observable; la flecha verde con dos puntas señala la distancia del horizonte. (Las escalas relativas utilizadas son sólo evocadoras; en realidad, las escalas difieren en factores demasiado grandes para incluirlas en un mismo dibujo.) En cada conjunto hay tres etapas de la evolución: un momento antes de la transición de fase (*arriba*), un momento después de la transición de fase (*el centro*) y hoy en día (*abajo*). En el modelo normal, la distancia del horizonte es siempre menor que la esfera gris, lo que dificulta la explicación de la uniformidad a gran escala observada en el universo. Puesto que en el modelo normal un dominio de la fase de simetría rota, creado en la transición de fase, tiene un radio comparable con la distancia del horizonte, deberían existir en el universo observado muchos monopolos y paredes de dominio. En el nuevo modelo inflacionario, la distancia del horizonte es siempre mucho mayor que la esfera gris, de tal modo que se espera que el universo observado sea uniforme a gran escala y tenga pocos (si los tiene) monopolos y paredes de dominio. En el modelo inflacionario, un instante antes de la transición de fase, la esfera gris es mucho menor que la del modelo normal; en la transición de fase, la esfera gris se expande en el modelo inflacionario multiplicando su radio por 10^{50} o más, para igualar el tamaño de la esfera del modelo normal.

es que de muchos de ellos se deducen transiciones de fase de caída lenta en rotación, sin ningún ajuste fino de los parámetros. La clave está en hallar un modelo supersimétrico que sea realista en lo que concierne a la física de partículas y que, además, también admita la inflación y el valor correcto para las inhomogeneidades de la densidad.

En resumen, el modelo inflacionario del universo es una teoría con pocas hipótesis que da cuenta de muchas características del universo observable, sin explicación en el modelo normal de la gran explosión. La belleza del modelo inflacionario reside en que, con él, la evolución del universo es casi independiente de los detalles de las condiciones iniciales, de las que poco o nada conocemos. Sin embargo, si el modelo inflacionario es correcto, será difícil descubrir consecuencias observables que se deriven de las condiciones que existían antes de la transición de fase inflacionaria. Análogamente, las grandes escalas de distancias creadas por la inflación hacen prácticamente imposible observar, en su totalidad, la estructura del universo. Lo que no obsta para que podamos estudiar estos aspectos, abordables desde distintos enfoques.

La descripción más sencilla del universo en momentos precoces del mismo es afirmar que empezó con una gran explosión, se expandió casi uniformemente hasta enfriarse a la temperatura crítica de la transición de fase y, a partir de aquí, siguió el modelo inflacionario. Extrapolando el modelo de la gran explosión hasta el tiempo cero, obtendríamos un universo reducido a una singularidad cosmológica, con una temperatura y densidad infinitas, al que no es aplicable ninguna de las leyes físicas conocidas. Una segunda posibilidad (como antes, sin explicación) es que el universo empezara con un estado caótico aleatorio. Las distribuciones de materia y temperatura serían no uniformes, con algunas regiones en expansión y otras en contracción. Dentro de este cuadro, ciertas regiones pequeñas, calientes y en expansión, sufrirían la inflación, evolucionando hacia inmensas regiones capaces de contener fácilmente el universo observable. Fuera de estas regiones, seguiría el caos e iría penetrando gradualmente dentro de las regiones que han sufrido la inflación.

Se ha especulado, desde planteamientos rigurosos, sobre si la creación real del universo podrían o no describirla las leyes físicas. La idea es que el universo habría surgido como una fluctuación cuántica a partir de la nada absoluta. Esta idea fue propuesta, en 1973, por Edward P. Tryon, del Hunter College de la Universidad de la Ciudad de Nueva York; en 1982, Alexander Vilenkin de la Universidad Tufts la reintrodujo, dentro del contexto del modelo inflacionario. En este contexto, "nada" puede referirse al espacio vacío, mas, para Vilenkin, representa un estado sin espacio, ni tiempo, ni materia. Las fluctuaciones cuánticas de la estructura del espacio-tiempo sólo pueden considerarse en el contexto de la gravedad cuántica; por consiguiente, estas ideas deben recluirse al ámbito de la especulación hasta que se formule una teoría operativa de la gravedad cuántica. Lo que no quita atractivo al hecho de que las leyes físicas puedan determinar no sólo la evolución del universo, a partir de un cierto estado, sino también las condiciones iniciales del universo observable.

El modelo inflacionario permite diversas posibilidades para la estructura global del universo. (Para todas ellas, el universo observable es sólo una parte muy pequeña del total; el límite de nuestro dominio estaría a unos 10^{35} años luz, si no más.) La primera posibilidad es que los dominios se hallen en contacto unos con otros y llenen todo el espacio. Los dominios están, entonces, separados por paredes de dominio y, en el interior de cada una, se mantendría la fase simétrica de la teoría de gran unificación. Los protones o neutrones que atravesaran tal pared se desintegrarían al instante. Las paredes de dominio tenderían con el tiempo a aplanarse. Transcurridos unos 10^{35} años, o más, los dominios más pequeños (incluido posiblemente el nuestro) desaparecerían, mientras crecerían los mayores.

Por otro lado, algunas versiones de las teorías de gran unificación no permiten una formación definida de paredes de dominio. En estas teorías puede ocurrir que estados diferentes de simetría rota, de dos dominios vecinos, se entremezclen suavemente. En la separación de ambos dominios, la densidad y velocidad de la materia tendrán discontinuidades; y se podría encontrar algún monopolo magnético.

Deduciríamos una posibilidad bastante distinta si la densidad de energía de los campos de Higgs fuera la que se indica en la parte inferior del dibujo de la figura 2. Como en los dos casos anteriores, habrá regiones del espacio que se sobreenfriarán en un estado de vacío falso, sufriendo una expansión acelerada. Como en el modelo inflacionario original, el falso vacío se desintegrará

mediante la formación aleatoria de burbujas: las fluctuaciones cuánticas provocarán que al menos uno de los campos de Higgs de una región pequeña del espacio atraviese, por efecto túnel, la barrera de energía, tomando el valor señalado en el dibujo con la letra *A*. En oposición a la versión inflacionaria original, a partir de ahí el campo de Higgs evoluciona muy lentamente (debido a lo plana que es la curva cerca de *A*) hacia su valor de vacío verdadero. La expansión acelerada sigue; bastaría una sola burbuja para albergar el universo observado. Si la intensidad de creación de burbujas es baja, escasearán las colisiones entre ellas. Las burbujas llenan una parte del espacio que estará cerca de la unidad cuando el sistema evolucione, pero el espacio se expandirá tan rápidamente que la parte de volumen que permanece en el estado de falso vacío aumentará con el tiempo. Eternamente continuarán formándose universos burbuja y no habrá manera de saber cuánto tiempo ha pasado antes de aparecer la burbuja. Esta descripción se parece mucho al viejo modelo cosmológico del estado estacionario a muy grandes escalas; lo que no empece que, dentro de cada burbuja, la evolución se acomode al modelo de la gran explosión, mejorado con la inflación.

Desde un punto de vista histórico, la idea del modelo inflacionario que quizá haya resultado más revolucionaria sea aquella según la cual toda la materia y la energía del universo observable puedan haber aparecido de casi la nada. Esta afirmación se enfrenta a la tradición científica multisecular para la que la nada no produce nada. Esta tradición, que se remonta, por lo menos, al siglo quinto antes de Cristo, con el filósofo griego Parménides, se ha concretado, en los tiempos modernos, con la formulación de cierto número de leyes de conservación, que establecen que ciertas cantidades físicas no pueden cambiar en ningún proceso físico. Diez años atrás, la lista de las cantidades que se creía se conservaban incluía la energía, el momento lineal, el momento angular, la carga eléctrica y el número bariónico.

Como el universo observado presenta un elevado número bariónico y una elevada energía, la inmensa mayoría de los físicos, salvo contadas excepciones, reputan insostenible la idea de creación a partir de la nada. (Las otras leyes de conservación que antes hemos mencionado no plantean este problema: el universo observado presenta un margen de valores para la carga eléctrica y el momento angular totales en el que cabe el valor cero, mientras que el momento lineal total depende de la velocidad del observador y no puede, por tanto, definirse en términos absolutos.) Sin embargo, con las teorías de gran unificación parece ahora completamente admisible que no se conserve el número bariónico. Sólo resta por considerar qué ocurre con la conservación de la energía.

La energía total de cualquier sistema puede dividirse en una parte gravitacional y una parte no gravitacional. La parte gravitacional (es decir, la energía propia del campo gravitacional), despreciable en un laboratorio, puede adquirir suma importancia en cosmología. La parte no gravitacional no se conserva por sí sola; en el modelo normal de la gran explosión, disminuye rápidamente cuando el universo primitivo se expande; la intensidad con la que pierde esta energía es proporcional a la presión del gas caliente. Por otro lado, durante la era inflacionaria, la región en estudio se llena con un falso vacío que tiene una gran presión negativa. En este caso, la energía no gravitacional aumenta drásticamente. Toda la energía no gravitacional del universo se crea, esencialmente, cuando el falso vacío sufre la expansión acelerada. Durante la transición de fase, esta energía se libera y termina por transformarse y engendra estrellas, planetas, seres humanos, etcétera. Según esto, el modelo inflacionario ofrece lo que parece ser la primera explicación científica plausible de la creación de prácticamente toda la materia y energía del universo observable.

En estas circunstancias, la parte gravitacional de la energía queda algo mal definida. Mas, a grandes rasgos, podemos decir que la energía gravitacional es negativa y que su valor compensa exactamente la energía no gravitacional. La energía total sería, cero y, por ello, compatible con la idea de la evolución del universo a partir de la nada.

Si las teorías de gran unificación aciertan en su predicción de la no conservación del número bariónico, no hay ninguna ley de conservación que se oponga a que el universo observado evolucione a partir de la nada. El modelo inflacionario del universo proporciona un posible mecanismo según el cual el universo observado podría haber evolucionado desde una región infinitesimal. Nada nos impide ceder a la tentación especulativa y dar un paso más: el universo ha evolucionado desde exactamente nada.

Un campamento de cazadores indios durante 20.000 años

Desde 19.000 a. C. hasta 1000 d. C. se ocupó periódicamente un abrigo rocoso del oeste de Pennsylvania. La extrema antigüedad del yacimiento afecta a la cuestión de cuándo penetró, por Bering, el hombre en el hemisferio occidental

J. M. Adovasio y R. C. Carlisle

¿Cuándo llegaron los primeros habitantes del Nuevo Mundo? Hasta hace poco, la opinión de los científicos estaba dividida en dos bandos principales. Algunos prehistoriadores suponían que las emigraciones de Asia a América del Norte empezaron hace 30.000 o 40.000 años, o incluso más. Otros, invocando la falta de pruebas seguras de una entrada tan antigua, situaban las emigraciones alrededor de hace 15.000 años, si no menos. Ahora, tras un decenio de trabajo de campo en un abrigo rocoso del oeste de Pennsylvania, todos los indicios sugieren que la respuesta se encuentra a medio camino entre ambos extremos.

El yacimiento, llamado Meadowcroft, está cerca del río Ohio, 75 kilómetros al oeste de Pittsburg. Se halla en terrenos de Albert Miller, un historiador local muy interesado en la arqueología; fue elegido para la investigación más por casualidad que por propósito deliberado. A principios de la década de los 70, uno de nosotros (Adovasio) andaba en busca de un sitio donde los estudiantes de arqueología de la Universidad de Pittsburg pudieran ejercitarse en técnicas de excavación. Un colega, el historiador Phil R. Jack, nos indicó que abordásemos a Miller para que nos permitiese trabajar en un abrigo de su tierra. Miller accedió gustoso y en junio de 1973 empezaron los trabajos. Aquí describiremos las excavaciones y explicaremos por qué las pruebas del uso sorprendentemente antiguo del abrigo dan un color completamente nuevo al cuadro del hombre primitivo en el Nuevo Mundo.

El abrigo rocoso de Meadowcroft se encuentra en la escarpada orilla norte de Cross Creek, un arroyo que discurre de este a oeste y desemboca en el Ohio a menos de trece kilómetros del yacimiento. En la actualidad, el viento dominante sopla del oeste y cruza la boca del abrigo, proporcionándole una ventilación casi constante. Creado por la erosión de pizarras situadas bajo arenisca de Morgantown-Connellsville, el abrigo tiene ahora 15 metros de ancho y seis de profundidad desde el borde de su techo hasta su pared trasera. En tiempos fue considerablemente mayor, según indican los muchos bloques de arenisca y piedras menores caídos en el piso actual del abrigo y debajo del mismo.

Antes de comenzar la excavación de Meadowcroft, nuestro equipo de campo hizo un minucioso inventario de las clases de plantas que crecían alrededor del abrigo, dentro de él, en la ladera de arriba y en la ladera de abajo. Hecho esto, toda la vegetación situada dentro de los veinte metros al este y al oeste del punto medio del abrigo fue arrancada, para poner al descubierto su actual topografía. La zona fue cartografiada y en ella se estableció un sistema de cuadrícula básica, compuesta de cuadrados de dos metros (cuando los trabajos posteriores necesitaron una cuadrícula más fina, los cuadrados fueron subdivididos en otros de un metro e incluso de 0,50 y 0,25 metros cuadrados). Un control permanente de altura fue adosado entonces a la pared trasera del abrigo, de modo que una serie de puntos descriptibles con precisión en tres dimensiones quedaron a disposición de los excavadores.

Tomadas estas medidas preliminares, empezó la excavación de la primera zanja. Iba de sur a norte a lo largo de unos doce metros, a partir de la ladera, por debajo de la actual línea de goteras (donde el agua que cae del techo del abrigo toca el suelo) y continuando hacia dentro. La excavación reveló la presencia de anteriores líneas de goteras y la secuencia de su retroceso, conforme partes del techo iban cayendo. Esto nos permitió calcular la cantidad de suelo protegido disponible en el interior del abrigo en fechas progresivamente anteriores.

Durante las varias campañas inmediatas, la ampliación de esta zanja inicial hacia el este y el oeste, así como la apertura de otras zanjas, revelaron, para sorpresa nuestra, que las capas de sedimentos que se habían acumulado en el suelo del abrigo eran mucho más gruesas que el metro aproximado que habíamos esperado. Su espesor total pasaba de los cinco metros y en él eran evidentes 11 claros estratos naturales. Dentro de cada uno de ellos, en los que faltaban microestratos naturales, los excavadores fueron levantando una tras otra, y a su arbitrio, capas de un centímetro, de cinco y de diez centímetros. Tanto los estratos principales como sus subdivisiones naturales estaban definidos por la contextura de los sedimentos, su grado de compacidad, su friabilidad (tendencia a desmenuzarse) y, en algún caso, diferencias de color. También observamos que estratos y microestratos de similar composición y grado de compacidad sonaban a menudo del mismo modo, al ser golpeados suavemente con la contera del mango de un palustre. Así, en combinación con más convencionales criterios de campo, hasta el sonido nos ayudó a seguir el curso de los distintos niveles de los depósitos del abrigo.

Asignamos números romanos a los once estratos, llamando al más alto Estrato XI y al más bajo, que no ofrecía señales de actividad humana, Estrato I. El estrato más bajo con muestras de habitación humana, el Estrato II, fue además subdividido en IIa y IIb, el primero de ellos el más bajo y más antiguo. Los

sedimentos de todos los estratos, excepto el Estrato IIa fueron tamizados en seco con una malla de seis milímetros, para averiguar su contenido, a saber, semillas, nueces, huesos y fragmentos de éstos. El estudio de tales restos de flora y fauna nos permite determinar cuál era el ambiente natural del yacimiento en distintas épocas del pasado y también averiguar lo que sus habitantes humanos estaban cazando, recolectando, comiendo y, en general, haciendo durante su estancia en el mismo. Los sedimentos recogidos en el Estrato IIa fueron tamizados, empapados de humedad, en una malla de 22 centímetros. Además de este sistema relativamente basto, se tomaron muestras de cada cuadrado excavado en el yacimiento (en un total de 147 metros cuadrados de superficie, tanto dentro como fuera de la línea de goteras moderna).

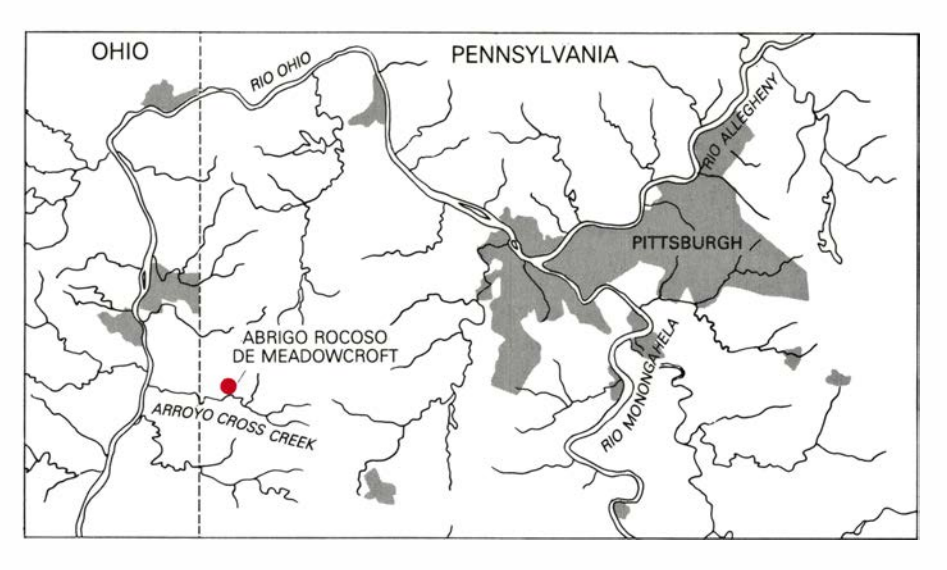
Estas muestras, representativas de las capas de un centímetro, cinco centímetros y diez centímetros que se iban rebajando en cada estrato del cuadrado, fueron tratadas de una de estas dos formas. Si la muestra representaba un relleno sin señales de alteración por el calor, o si procedía de un ambiente cultural distinto del hoyo de una hoguera o de un hogar (verbigracia, una fosa de basura o de almacenamiento), se sumergía en agua y después se pasaba por tamices graduados (el tamiz más fino tenía una malla de 74 micromilímetros). Si la muestra era de una hoguera o de un hogar, la flotación no se hacía con agua, sino con peróxido de hidrógeno. Esto reducía al mínimo el daño a las semillas carbonizadas, nueces o fragmentos de madera que la muestra pudiese contener. En conjunto, la excavación recogió un total de 230 metros cúbicos de sedimentos para su examen.

En estos totales se incluyen los resultados de doce "columnas" de muestreo. Estas fueron cortadas en las caras descubiertas de las zanjas y sometidas a análisis geológicos y de otras clases. Las columnas llegaban desde la superficie del yacimiento hasta el Estrato I, y en dos casos penetraban a fondo en este estrato. De las columnas se extrajeron muestras en bruto a intervalos de cinco a diez centímetros. Cada muestra fue troceada y las fracciones resultantes analizadas para determinar variables tales como la proporción de fango a arcilla y a partículas del tamaño del grano de arena o mayores; así como para averiguar la cantidad de carbonato de calcio en el material y la abundancia relativa de elementos como meros indicios en cada nivel. Además de estas pruebas geológicas, las fracciones fueron examinadas en busca de restos microscópicos vegetales y animales.

Otra investigación geológica consistió en la recogida de muestras de roca de la superficie misma del peñasco. Las tomas se efectuaron a intervalos de 20 a 25 centímetros, desde la base de la arenisca de Morgantown-Connellsville, dentro de la cual se formó el abrigo, hasta la cima de esta formación geológica, en una distancia de unos 20 metros. Las muestras fueron empleadas para cortar placas delgadas. La composición y el tamaño del grano de las varias placas fueron entonces comparadas con similares placas-muestra hechas de la arenisca que había caído del techo del abrigo, con objeto de determinar el esquema y el



1. EXCAVACION EN PROFUNDIDAD, en el abrigo rocoso de Meadowcroft. Las paredes están marcadas con etiquetas blancas, que señalan los distintos estratos del yacimiento; sitúan elementos tales como hogares, silos de almacenamiento y zonas de trabajo, e identifican puntos en que se han tomado muestras para análisis de polen y otros fines. Los excavadores alcanzaron el suelo virgen (Estrato I) a más de cinco metros. Descubrieron que diez estratos superpuestos contenían las pruebas de las visitas que hicieron al yacimiento las gentes de tres culturas: paleoindia, arcaica y de Woodland.



2. ABRIGO ROCOSO DE MEADOWCROFT, en la ribera norte de Cross Creek, a unos 75 kilómetros de Pittsburgh. Estudiantes de arqueología empezaron a excavar en dicho lugar en el verano de 1973.

mecanismo de la erosión que había dado forma al abrigo.

La mayor parte de la secuencia prehistórica de Meadowcroft, desde lo alto del Estrato II hasta lo alto del Estrato IX, pertenece al período cultural que los arqueólogos norteamericanos llaman de Woodland ("de bosque"). Una larga serie de cálculos de edad, mediante el carbono-14, indican que este período abarcó entre 2250 y 2500 años aproximadamente en este yacimiento, empezando alrededor de 1115 a. C. Para describir los hallazgos más significativos del Woodland realizados en este abrigo conviene que invirtamos el orden de su descubrimiento y nos movamos del pasado al presente. De este modo, los restos más antiguos del Woodland vienen de la parte alta del Estrato III, donde se encontró un gran hoyo de una hoguera con tiestos de varios vasos cerámicos. Las cerámicas, hechas utilizando mineral de hierro pulverizado como degrasante, y decoradas con impresiones de cuerdas, son el ejemplo más antiguo de cerámica india encontrado en la parte alta del valle del río Ohio. Las semillas conservadas de calabaza de la misma época, esto es, hacia 1115 a. C., indican que la horticultura, que es un rasgo clave de la cultura de Woodland, llegó probablemente a esta parte del alto Ohio más de 500 años antes de lo que se había supuesto.

La continuidad de la horticultura está indicada por la presencia de una semilla de calabaza en la parte baja del estrato sobrepuesto, el Estrato IV, que data de hacia 865 a. C. El hallazgo de un carozo carbonizado de una espiga de 16 hileras, de la parte superior del mismo estrato, que data de 375 a 340 a. C., señala una nueva fecha temprana para el cultivo del maíz en la zona del alto Ohio.

La llegada de la horticultura coincide aproximadamente con el fin de la utilización relativamente intensiva de Meadowcroft como abrigo. Los alrededores inmediatos del yacimiento carecen de las amplias llanuras inundables que eran zonas naturales para jardines y emplazamientos de aldeas entre las gentes del Woodland. El abrigo continuó siendo un refugio de cazadores y una base para la recolección y preparación de alimentos silvestres, pero las señales de tales actividades son por lo general escasas. La tendencia al abandono del abrigo fue acelerada sin duda por el aumento de las caídas de rocas del techo, en su extremo oriental, entre 300 y 600 d. C. El derrumbamiento redujo por lo menos en un 20 por ciento el espacio de suelo disponible. Por el tiempo en que los primeros colonos europeos entraban en contacto con la población indígena de esta parte de Pensylvania, a principios del siglo XVIII, la zona de Meadowcroft parece haber estado desierta.

La utilización del abrigo de Meadowcroft por las gentes del Woodland se vio precedida por unos 7000 años de visitas de los cazadores y recolectores de una cultura india anterior: la Arcaica. Moviéndonos otra vez de los niveles más antiguos a los más recientes, los Indios Arcaicos dejaron huellas de su presencia desde la parte alta del Estrato IIa, que data de entre 8500 y 8000 a. C., hasta debajo justamente de la parte alta del Estrato III, fechado hacia finales del segundo milenio a. C. El comienzo del período Arcaico en Meadowcroft coincide aproximadamente con la retirada de la última capa de hielo glacial de América del Norte y el comienzo del Holoceno: la época geológica presente.

Desde entonces y hasta la llegada de las gentes del Woodland, el abrigo y sus alrededores volvieron a ser visitados periódicamente por grupos de indios arcaicos que practicaban una economía de subsistencia como cazadores y recolectores. Parece seguro que Meadowcroft albergó más visitantes humanos a medida que el período Arcaico avanzaba, pero no está del todo claro si esto supone grupos mayores o un mayor número de grupos pequeños, o simplemente visitas más largas o más frecuentes.

Comoquiera que sea, estas gentes perseguían por temporadas a una fauna abundante: ciervos, alces, pavos silvestres y una variada caza menor. Las gentes del Arcaico también explotaban los recursos de la flora de la zona, como lo acredita la presencia en fosas de almacenaje y de desecho, entre otros restos de plantas, fragmentos de cáscaras de nuez y las semillas y otros restos de almezas y de almezas enanas, a millares en dichas fosas. A los cazadores del Arcaico hay que atribuir también una extensa variedad de puntas de piedra para proyectiles, otros utensilios de piedra y herramientas de hueso y de madera. En rigor puede afirmarse que Meadowcroft nunca fue más intensamente utilizado que a fines del período Arcaico.

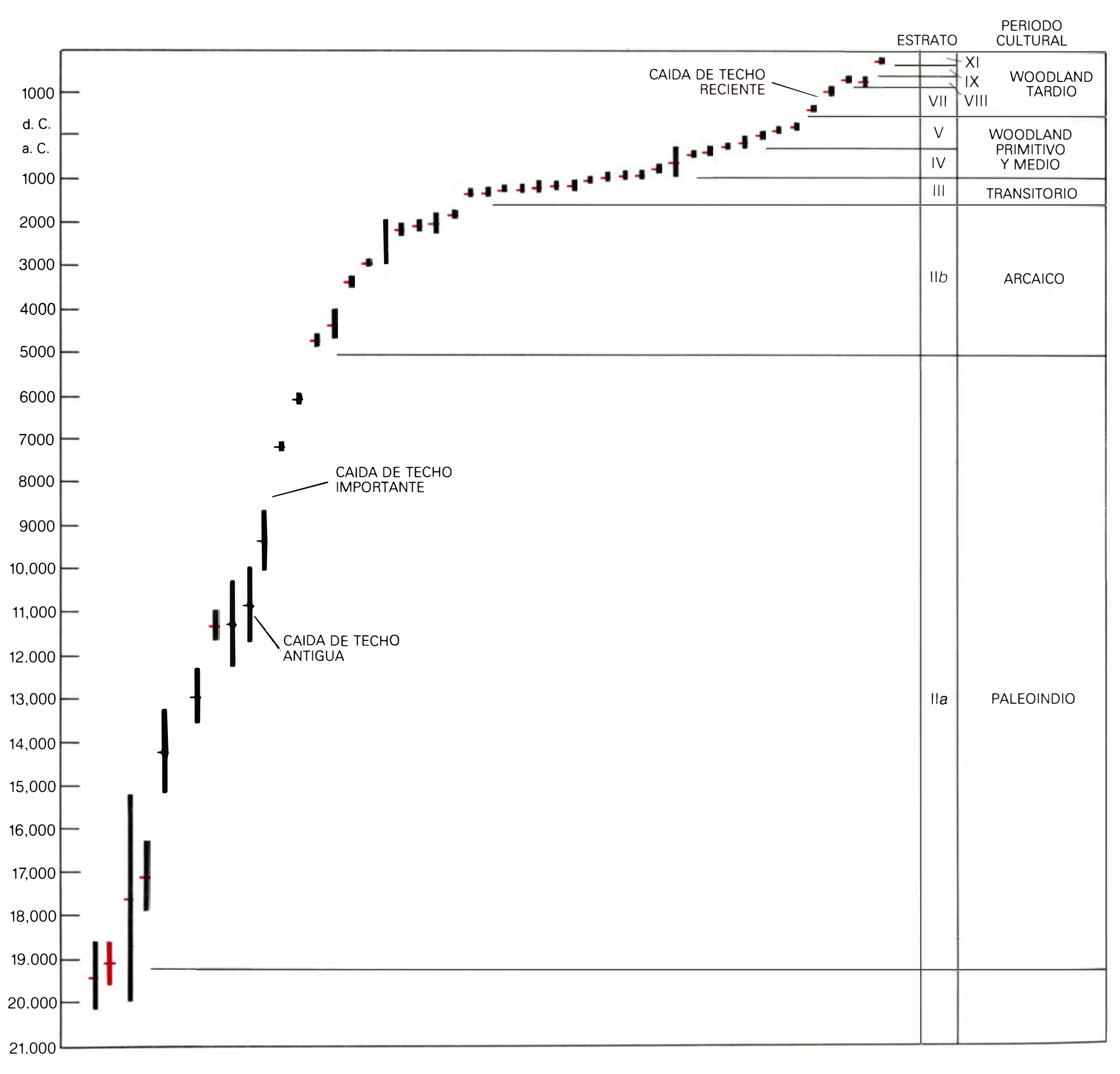
El haber descubierto una secuencia continua de ocupación prehistórica, de una duración de más de diez milenios, en un yacimiento seleccionado en principio para fines educativos, fue motivo de profunda satisfacción. Nadie esperaba en serio que la parte del Estrato IIa situada entre los niveles del Arcaico primitivo, por arriba, y el Estrato I estéril, por abajo, encerrase sorpresas de ningún género. En todo caso concluimos la campaña de 1973 enviando la primera de unas cien muestras de carbón, que abarcaban la secuencia completa de los depósitos del abrigo, a Robert Stuckenrath, del laboratorio de radiobiología de la Smithsonian Institution, para la datación por carbono-14.

Las primeras fechas estuvieron disponibles en el invierno de 1973-74. Correspondían a muestras que Stuckenrath asignaba al XII milenio a. C. Al principio contemplamos aquellas fechas tan altas con no poco escepticismo. En aquella época tanto nosotros como otros prehistoriadores de tendencia conservadora del Nuevo Mundo sustentábamos la opinión de que el

hombre no había llegado a las Américas hasta casi el final del período Pleistoceno, hace unos 15.000 años.

Otras muestras del yacimiento fueron enviadas a la Smithsonian para su datación al final de la campaña de 1974, y otras más en el curso de los dos años siguientes. A la hora de recibir los resultados finales de la datación, en el invierno de 1978, ninguno de los relacionados con el estudio de Meadowcroft seguía teniendo reserva alguna acerca de la edad relativamente elevada del Estrato Inferior IIa. Desde el fondo hasta la cima la secuencia de fechas del carbono-14 no sólo tenía coherencia interna, sino que también la tenía con la estratigrafía del yacimiento. Las fechas más remotas oscilaban ahora entre 9350 ± 700 a. C. (Estrato Medio IIa) y 17.650 ± 2400 a. C. (Estrato Inferior IIa). En fecha reciente un laboratorio comercial de datación radiactiva nos dio una determinación de carbono-14 aún más antigua. La muestra reveló la fecha de 19.120 ± 475 a. C. Estaba sacada del nivel más bajo del Estrato IIa, pero no está asociada, al parecer, con actividades humanas.

Los habitantes más primitivos del hemisferio occidental se asignan a un estadio cultural definido grosso modo como Paleoindio. Sus características más familiares son la caza de animales de caza mayor (hoy extinguidos en su mayor parte) y la fabricación de puntas de proyectiles de piedra elegantemente talladas. La comprobación de que los paleoindios habían llegado a Meadowcroft, en el este de América del Norte, hace más de 19.000 años levantó una enorme polvareda entre los arqueólogos americanos. El quid de la controversia está precisamente en cuándo los primeros emigrantes de Asia cruzaron el puente de tierra de lo que es ahora el Estrecho de Bering. Si algunos de ellos habían alcanzado la cuenca alta del río Ohio hacia 17.000 a. C., parece lógico que, dándoles algún tiempo para su emigración hacia el este, hubiesen cru-



3. FECHAS DE CARBONO-14 obtenidas de muestras de carbón excavadas en el abrigo. Las más antiguas aparecen a la izquierda y las más recientes a la derecha. El guión de color indica la fecha media en cada caso; la longitud de la barra vertical señala los valores de más-menos. Todas las dataciones, excepto la inferior izquierda (*color*), fueron establecidas por el laboratorio de radiobiología de la Smithsonian Institution, que analizó las muestras.

zado el istmo por lo menos 2000 o 3000 años antes. Pero resulta que tal fecha de entrada es miles de años anterior a la fecha aproximada que se suele atribuir a ese suceso: 13.000 a. C. o más tarde. Esto no les dejaba más que una alternativa a los que se aferraban a la opinión acostumbrada, a saber: las fechas altas de Meadowcroft deben de estar equivocadas, quizá porque las muestras de carbón estaban contaminadas.

¿Cómo podía ser eso? Una de las explicaciones era que los depósitos de carbón situados a un kilómetro del yacimiento producían "carbón antiguo": una substancia particulada que había contaminado las muestras de carbón de Meadowcroft haciéndolas parecer más antiguas de lo que eran realmente. Cabía también la alternativa de que las aguas freáticas del yacimiento pudieran ser ricas en "carbón antiguo" disuelto, procedente del depósito de carbón, lo que habría producido el mismo efecto en las muestras.

De hecho, como parte del procedimiento normal de pretratamiento en el laboratorio de la Smithsonian, cada una de las muestras de Meadowcroft había sido examinada al microscopio en busca de una posible contaminación particulada. Ninguna de las muestras acusaba señal alguna de tal contaminación. En cuanto a la posibilidad de contaminación por "carbón antiguo" disuelto, suministramos a Stuckenrath muestras de madera del período, rico en carbón, que en general se conoce como Carbonífero (en el norteamericano usual, Pensilvanio), sacadas de las proximidades del abrigo. Esta era la única fuente remotamente posible de contaminantes de carbón antiguo, en solución, en alguna parte cerca del yacimiento. Stuckenrath fue incapaz de disolver las muestras, ni hirviéndolas en agua ni en otros varios disolventes. Queda así de manifiesto que ninguna de las fuentes propuestas de posible contaminación intervino en Meadowcroft.

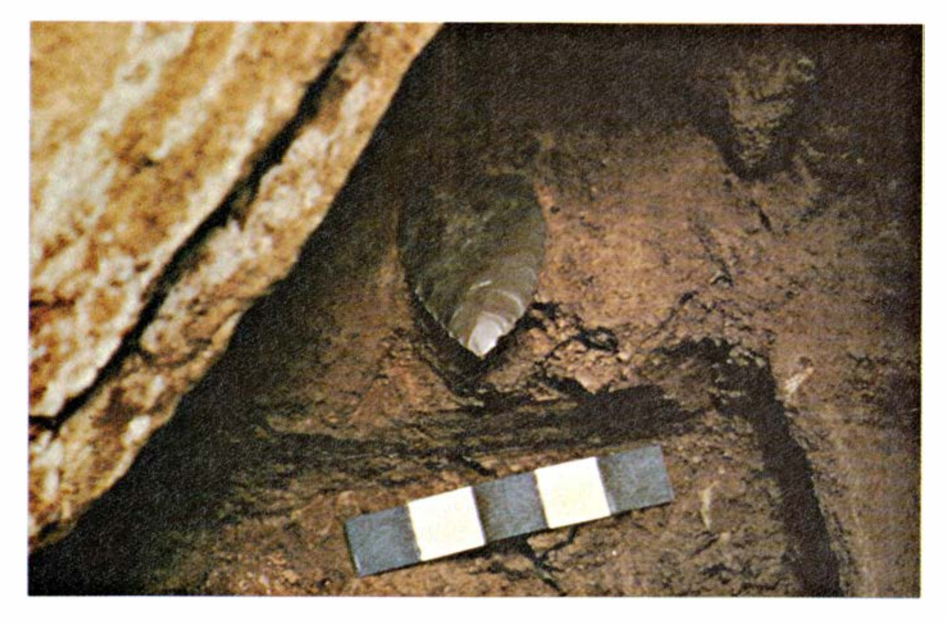
Quienes critican las fechas más antiguas del yacimiento parecen haber pasado por alto otro argumento muy sólido en contra de su postura. Es seguro que cualquier mecanismo de contaminación que suponga la circulación de agua subterránea no podía haber seleccionado, para afectarlas únicamente a ellas, las muestras más antiguas del yacimiento. Y, sin embargo, las fechas más recientes de este yacimiento –las del Paleoindio terminal, y las de los niveles Arcaico y de Woodland– nunca han sido discutidas. Al contrario, como hemos observado, son extraordinariamente firmes.

Una última palabra sobre el asunto. Hay que subrayar que la única muestra del nivel profundo que fue datada por un laboratorio comercial se entregó "a ciegas". No se facilitó ni el nombre del yacimiento ni la posición estratigráfica de la muestra (el Estrato IIa). Al laboratorio sólo se le dijo que la muestra procedía de la zona general de Cross Creek y que probablemente tenía más de nueve mil años de antigüedad. Los sistemas de procesado y de recuento fueron distintos de los empleados por el laboratorio de la Smithsonian. La muestra resultó ser la más antigua de todas las del yacimiento (19.120 ± 475 a. C.), lo que está de acuerdo con la gran profundidad a que se encontraba en los depósitos del abrigo.

¿Cuáles son los restos culturales más antiguos recuperados en Meadowcroft? Para responder a esa pregunta conviene describir el Estrato IIa en detalle y sus varias fechas. En general, el estrato se compone de sedimentos procedentes, en su mayor parte, de la erosión del techo y de las paredes del abrigo. Se le puede dividir en tres subunidades de espesor desigual. En la parte superior de la subunidad más alta, Estrato Superior IIa, hay un suelo de ocupación humana que proporcionó carbón que data de 6060 ± 110 a. C. Un piso de ocupación un poco más profundo, dentro de la misma subunidad, data de 7125 ± 115 a. C. En la base de la subunidad hay una cantidad considerable de roca caída del techo que señala el límite entre la subunidad superior (Arcaico primitivo) y el Estrato Medio IIa (Paleoindio). Inmediatamente por debajo de la roca caída del techo, otro suelo de ocupación proporcionó una fecha más antigua aún: 9350 ± 700 a. C. No se han calculado más fechas para el resto de la segunda subunidad; su fondo se superpone a otra capa de piedras caídas del techo. Una muestra de un suelo de ocupación situado inmediatamente por debajo de esta capa, en lo alto del Estrato Inferior IIa, data de 10.850 ± 870 d. C. Siete muestras todavía más bajas de esta subunidad, la más profunda de todas, alcanzan una edad que frisa en los 11.290 ± 1010 y los 17.650 ± 2400.

Para redondear estas fechas puede decirse que el Estrato Superior IIa data de 9000 a 6000 a. C. y pertenece, por tanto, al Holoceno en su totalidad. El Estrato Medio IIa, por su parte, puede decirse que data de 11.000 a 9000 a. C. y, por ende, equivale a la terminación del Pleistoceno. Por último, las partes del Estrato Inferior IIa que dan señales de ocupación humana puede decirse que datan de 17.500 a 11.000.

El testimonio más antiguo del uso de instrumentos de piedra se compone de 123 útiles completos o fragmentados y de varios cientos de lascas de desperdicio de la fabricación o el mantenimiento de aquéllos. Todos ellos estaban asociados con elementos de suelos de ocupación del Estrato IIa Medio e Inferior. Los había bifaciales (tallados por



4. PUNTA DE PROYECTIL vista *in situ*, parcialmente liberada de la matriz arenosa que la rodea, encima de la escala en centímetros que la acompaña en esta fotografía. Los excavadores han denominado a esta punta y a otras semejantes halladas en la zona, Lanceolada Miller, en honor de Albert Miller, dueño del terreno en que está situado Meadowcroft. Es una de las más antiguas puntas paleoindias descubiertas hasta ahora al este del río Mississippi: se fecha entre 10.850 ± 870 y 9350 ± 700 a. de C.

ambas caras). Algunos cuchillos sólo estaban tallados por un lado. Entre ellos hay una forma característica de cuchillo, el "mungai", llamado así por otro yacimiento de Cross Creek, Mungai Farm. También había hojas, lascas de varias formas, que habían sido utilizadas con distintos fines y fragmentos de los "núcleos" de los cuales se habían extraído los utensilios.

El suelo de habitación de la parte alta del Estrato Medio IIa proporcionó la primera y más antigua punta de proyectil encontrada en Meadowcroft. Hemos bautizado el tipo con el nombre de Lanceolado Miller, en honor de nuestro anfitrión de Meadowcroft, Albert Miller. Otras puntas de proyectil del mismo tipo han sido halladas en el yacimiento de Mungai Farm, en el yacimiento de Pershina (que volveremos a investigar este verano) y en otros lugares dentro o en las cercanías de la cuenca de Cross Creek. Hasta ahora, estos yacimientos no han proporcionado puntas del Lanceolado Miller en ningún contexto estratigráfico que permitiese fecharlas directamente. El Lanceolado Miller de nuestro Medio IIa, sin embargo, es una de las más antiguas puntas de proyectil exhumadas en el este de América del Norte. Puede ser el antepasado local de las puntas paleoindias de amplia difusión, llamadas Clovis.

Estos niveles inferiores de Meadowcroft se hallan fuera de la actual línea de goteras del abrigo, de modo que el material orgánico que hubiera en ellos estaba mal conservado. Afortunadamente, se recuperó bastante material para dar un cuadro general del ambiente de la época en que los paleoindios primitivos visitaron la comarca. Por ejemplo, el polen y otros restos vegetales indican que tanto árboles de hoja caduca como coníferas crecían en la vecindad. Por lo que se refiere a los animales, los cazadores, muy al principio, tenían a su disposición ciervos de cola blanca y caza menor, v.gr. ardillas (*chipmunks*) y ardillas voladoras del sur. Es muy improbable que no hubiese otros animales en la zona de Meadowcroft, pero ninguno ha dejado huellas identificables en los niveles inferiores del yacimiento.

Por apasionante que sea el haber constatado la presencia de bandas de cazadores en el oeste de Pennsylvania hace de 16.000 a 19.000 años, la verdadera importancia del Proyecto Arqueológico de Meadowcroft-Cross Creek radica en algo más que el contenido de los estratos más profundos del abrigo. Además de descubrir los niveles Arcaico y de Woodland del abrigo, nuestra obra interdisciplinar procuró abarcar toda la cuenca de Cross Creek, un área de unos ochenta kilómetros cuadrados, a partir de la confluencia del arroyo con el río Ohio y extendiéndose al este y sudeste hasta la cabecera del arroyo. Entre 1973 y 1978 los arqueólogos de campo registraron 231 yacimientos prehistóricos en la cuenca. De ellos, 17 fueron extensamente explorados y dos enteramente excavados, yacimientos que encierran artefactos paleoindios, arcaicos y de Woodland.

Los arqueólogos del Proyecto de Meadowcroft-Cross Creek han reunido unos 20.000 artefactos de piedra, tanto tallada como pulimentada, amén de útiles hechos de hueso y de otros materiales perecederos. Los zoólogos del Proyecto han reunido casi un millón de huesos de animales y los botánicos recogido 1.400.000 especímenes de plantas. Las muestras geológicas se cuentan por millares. Los datos de todas estas colecciones se han preparado, o están preparándose, para análisis por computador y ofrecen una oportunidad simpar para examinar la adaptación de diversas poblaciones humanas a la vida en el sistema de una sola cuenca, durante casi veinte milenios.

Las cuestiones a que cabe responder antes de que el trabajo del Proyecto se dé por concluso son muchas y delicadas. Por poner un ejemplo: teniendo que ver con el Arcaico terminal, uno puede preguntarse qué efectos ejerció el clima postpleistoceno en los pobladores de la zona de Cross Creek. ¿Fue acaso este cambio climático responsable del aumento del número de yacimientos arcaicos y de la aparición de campamentos-base en el abrigo y en otros lugares? Otro ejemplo: ¿se corresponde la decreciente colección de plantas silvestres del período de Woodland con la llegada del cultivo de la calabaza y del maíz? ¿Qué factores explican la posterior desaparición de los campamentos-base de Meadowcroft y del área de Cross Creek en general? Aunque nuestro proyecto no hubiese conseguido otra cosa, podría exhibirse como un esfuerzo, coronado de un éxito sin precedentes, para crear un banco de datos interdisciplinar, un banco que haga posible la realización de un estudio intensivo de la ecología de una sola región durante un lapso muy largo de tiempo prehistórico. Hay buenas razones para esperar que esta riqueza de información se revele útil para prehistoriadores y estudiosos de otras disciplinas.

Gangliósidos

A pesar de su complejidad estructural, se conoce ya la composición química de 40 de ellos, así como los mecanismos de su formación y degradación. Intervienen en importantes funciones biológicas

José A. Cabezas y Pedro Calvo

Para muchos, el término gangliósido va asociado a una terrible enfermedad de la infancia, la idiocia amaurótica familiar, forma de imbecilidad que se caracteriza, además, por la pérdida de la visión y que resulta mortal. Menos conocida es, sin embargo, su destacada participación en uno de los mecanismos implicados en la pervivencia del ser humano: la transmisión del impulso nervioso.

Si fijamos nuestra atención en la composición química del sistema nervioso humano, advertimos que posee una pequeña proporción de glúcidos, en particular galactosa, glucosa y glucógeno. Más elevada es la concentración de proteínas en el mismo. Parte de ellas se encuentran unidas a algunos azúcares sencillos (monosacáridos), formando glicoproteínas. La concentración de lípidos en el sistema nervioso es bastante elevada, sobre todo en forma de colesterol y lípidos complejos; menos altas, las concentraciones de glicéridos. El funcionamiento adecuado del sistema nervioso depende muy directamente de otras sustancias químicas de una clase especial de lípidos complejos llamados esfingolípidos. Nombre que les viene de poseer una base esfingoide, que suele ser la esfingosina. El prefijo esfingo lo introdujo, en recuerdo de los enigmas atribuidos por los antiguos a la esfinge, Thudichum, descubridor de los cerebrósidos, a finales del siglo XIX.

Los gangliósidos, que se cuentan entre los esfingolípidos, se hallan en todos los tejidos animales, desde los equinodermos hasta los mamíferos, si bien en proporciones muy variables. En un comienzo se les asignó una ubicación restringida en los órganos neurales, pero con el progreso de las técnicas analíticas se les fue detectando en otras partes. Se ha calculado que aproximadamente un 6 por ciento de los lípidos de la sustancia gris cerebral corresponde a gangliósidos; esto es, diez veces más que la sustancia blanca. Se ha medido asimismo su concentración en la médula y la corteza adrenales, bazo, placenta, hígado, riñones, pulmón, intestino, plasma sanguíneo, eritrocitos y líquido cefalorraquídeo. No se han encontrado en el reino vegetal. En 1983, H. Rahmann y colaboradores, en Alemania, exponían sus resultados sobre la filogenia de los glangliósidos cerebrales en distintas especies de vertebrados.

Acerca de la distribución subcelular de los gangliósidos, señalemos que se hallan en la capa externa de todas las membranas plasmáticas, con la porción glucídica, hidrófila, orientada al exterior. (Datos experimentales recientes indican que esa porción, similar o idéntica a la de algunas glicoproteínas, se encuentra más próxima a la superficie externa de la membrana citoplasmática que la respectiva porción glucídica de dichas glicoproteínas.) Elevadas concentraciones de gangliósidos se albergan en las fracciones microsómica y sinaptosómica; en menor proporción, en el aparato de Golgi.

Naturaleza química y antecedentes

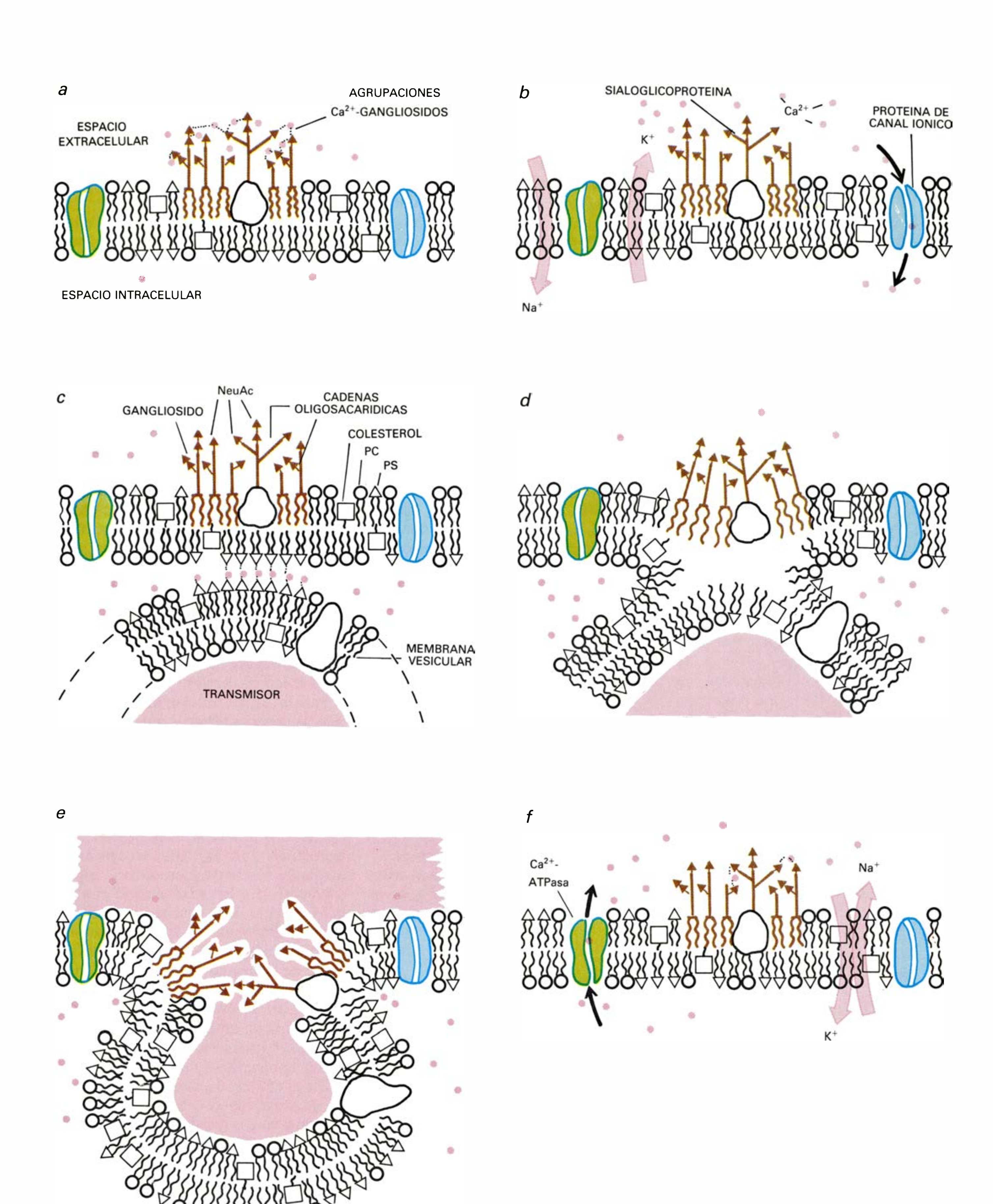
Hemos dicho que los gangliósidos son esfingolípidos. Estos se dividen en tres clases: ceramidas, esfingomielinas y glicoesfingolípidos. Las ceramidas resultan de la unión de la esfingosina con un ácido graso mediante un enlace de tipo amídico; son, pues, esfingoides *N*-acilados. En la segunda clase, la de las esfingomielinas, se une un componente de fosfocolina al esfingoide *N*-acilado, resultando la ceramida-1-fosfo-colina. Las esfingomielinas son prácticamente insolubles en éter; se diferencian de los restantes esfingolípidos por contener fósforo. Finalmente, los glicoesfingolípidos corresponden a lípidos compuestos por un esfingoide y, por lo menos, un monosacárido.

Los glicoesfingolípidos pueden dividirse en dos grandes grupos: neutros (cerebrósidos) y ácidos. El grupo de los ácidos está formado por los gangliósidos y los sulfoglicosilesfingolípidos.

Los gangliósidos o sialoglicoesfingolípidos fueron descubiertos por E. Klenk, en Colonia, hace unos 50 años, quien pensó inicialmente que se trataba de un único gangliósido y lo denominó "Gangliosid", por haberlo extraído de células *gangli*onares (células nerviosas grandes) y contener moléculas glucídicas (*os*as). Hacia 1970 se conocía ya la estructura de una veintena de ellos; en la actualidad, se conoce al menos parcialmente la estructura de unos 40.

En la investigación sobre los glicoesfingolípidos se ha pasado, de manera discontinua, por varias fases o etapas: aislamiento de los mismos a partir de órganos neurales y su caracterización a partir de los materiales donde se acumulan anormalmente (en pacientes con trastornos denominados esfingolipidosis); extracción y caracterización a partir de órganos neurales normales y extraneurales, donde se localizan, en menor proporción, tanto en seres humanos como en animales; vías metabólicas de biosíntesis y degradación de dichos esfingolípidos, cuyas deficiencias explicarían su acúmulo en circunstancias patológicas, y cuyo funcionamiento equilibrado se manifiesta como un patrón normal para cada especie y órgano o tejido; funciones biológicas, con sus mecanismos correspondientes; y aspecto genético de los procesos metabólicos implicados, así como intentos de corrección de las anomalías.

Los gangliósidos más sencillos son los formados por ceramida-glucosa-ácido *N*-acetilneuramínico (abreviadamente, Cer-glc-NeuAc) o por ceramida-galactosa-ácido *N*-acetil-neuramínico (Cer-gal-NeuAc). Para el ácido siálico la configuración es α, mientras que para la *N*-acetilhexosamina y la glucosa es β; para la galactosa suele ser también β, pero se conocen algunos gangliósidos en que es α. Aun tomando en consideración solamente la porción oligosacarídica, la complejidad de algunos gangliósidos es considerable. En lo que se refiere a la naturaleza de los ácidos



1. ESTADOS FUNCIONALES DE LA MEMBRANA presináptica durante la neurotransmisión. En el estado de potencial de reposo (*a*), las agrupaciones Ca^{2+}-gangliósidos inducen rigidez local. Bajo el potencial de acción (*b*), el Ca^{2+} se separa de los gangliósidos y entra a través de canales iónicos. El área de agrupación de gangliósidos se hace más fluida y más permeable a compuestos polares. Los tres esquemas siguientes (*c*, *d*, y *e*) ilustran la unión de las vesículas sinápticas, fusión y liberación del neurotransmisor; los iones Ca^{2+} forman en el interior complejos cruzados entre las vesículas y la parte interna de la membrana presináptica, después de haber inducido una separación de fases con formación de agrupamientos de fosfolípidos con carga negativa (fosfatidilserina = PS). En el estado de repolarización (*f*), por fin, el balance iónico inicial se restituye, el Ca^{2+} se vuelve a unir a los gangliósidos en el exterior, después de haber sido sacado por transporte activo. (Reelaborado de un esquema de H. Rahmann et al., J. Exp. Med. 52, 6, p. 282, 1982.)

grasos que integran los distintos gangliósidos, predominan los de C_{14} a C_{18} en ciertos materiales, y los de C_{20} a C_{26} en otros.

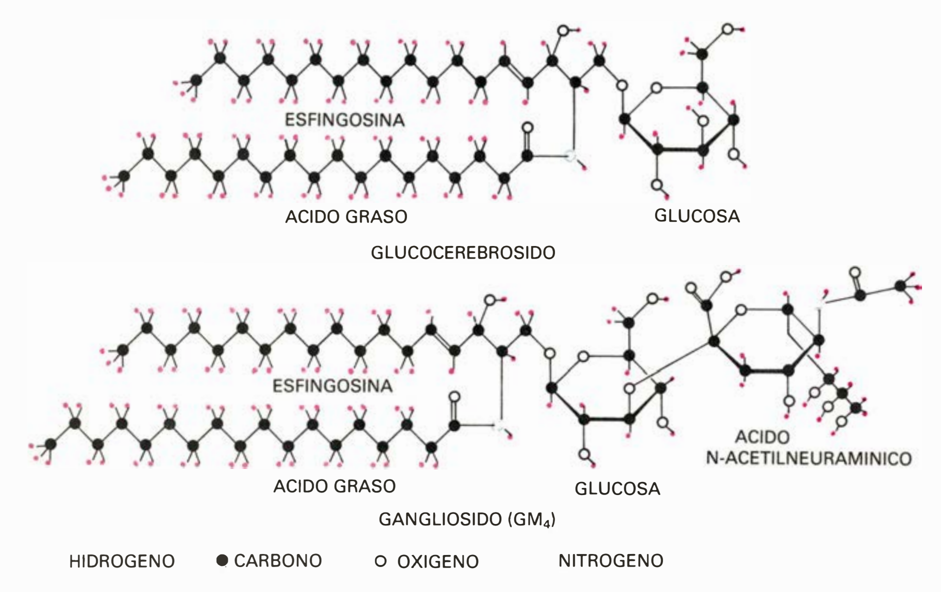
Nomenclatura

A medida que se fueron descubriendo nuevos gangliósidos se emplearon distintas denominaciones, como las usadas por Klenk, por Korey y Gonatas, por Kuhn y Wiegandt, y por Svennerholm, que empezaron a proliferar desde 1963. De todas ellas, la propuesta por Svennerholm, modificada más tarde por él mismo, es la que se ha impuesto, por su triple ventaja: es sencilla, mnemotécnica y racional. Emplea la letra *G* como designador común de todos los glangliósidos, seguida de *M, D, T* o *Q*, según sean uno (mono), dos (di), tres (tri) o cuatro (tetra), respectivamente, los restos de ácido siálico; así, se denominan monosialogangliósidos, disialogangliósidos, etcétera, los compuestos correspondientes. Añade como subíndice el número 1 cuando la cadena de osas-osamina es larga, o sea, de cuatro restos; el número 2 cuando esta cadena carece de galactosa; el 3, cuando faltan la galactosa y la *N*-acetilgalactosamina. Finalmente, se ponen a veces las letras *a* o *b* para distinguir entre sí aquellos gangliósidos que sólo difieren por las posiciones de los ácidos siálicos.

Andando el tiempo, a medida que se iban estableciendo las estructuras de otros gangliósidos, resultaba insuficiente dicha nomenclatura. En efecto, la cuarentena de gangliósidos hasta ahora conocidos difieren entre sí por varios motivos: *a*) el tipo de ácido siálico (siendo frecuentemente el *N*-acetilneuramínico, pero pudiendo ser el *N*-glicolilneuramínico u otros); *b*) el número de restos o residuos de ácidos siálicos, desde uno hasta siete; *c*) las posiciones de dichos residuos de ácidos siálicos en la molécula del gangliósido, cuando éstos son dos o más de dos; *d*) el tipo de acetilhexosamina (generalmente *N*-acetilgalactosamina, aunque en ciertos casos *N*-acetilglucosamina); y *e*) la longitud de la cadena oligosacarídica restante.

Con objeto de expresar adecuadamente la estructura de los lípidos, las recomendaciones de 1976 de la IUPAC-IUB (International Union of Pure and Applied Chemistry, International Union of Biochemistry) introdujeron una terminología más detallada y precisa, pero menos fácil de retener. Habida cuenta de que los glicoesfingolípidos neutros, precursores estructurales de los gangliósidos, pueden ser adscritos a varias series o familias, se ha establecido que los gangliósidos que posean glucosa queden adscritos a las familias "lacto" "neolacto" y "ganglio"; el gangliósido que posee galactosa en vez de glucosa se inscribe en la serie "gala". Se acepta, asimismo, la simplificación de los nombres de la cadena oligosacarídica.

La posición del resto glucídico respecto al extremo de la ceramida se señala con números romanos: I remite al constituyente inmediato; II, III o IV designará dicha posición para los siguientes. Por último, la posición del enlace correspondiente a cada resto respecto a la que se une del otro residuo se marca con un número arábigo, a modo de exponente del número romano antes mencionado. De forma esquemática, se representa mediante pequeñas figuras geométricas (por ejemplo, con cuadrados para la galactosa, triángulos para los ácidos siálicos, etcétera) la naturaleza de los constituyentes glucídicos.



2. ESTRUCTURA de un glucocerebrósido y de un gangliósido (*GM*₄). En los gangliósidos siempre está presente algún resto de ácido siálico. Se conoce ya la composición química de una cuarentena de ellos.

Métodos

Nos limitaremos a los métodos de separación y caracterización de gangliósidos. Klenk empleó los propios de la química orgánica; Folch-Pi extraía los lípidos de un tejido mediante mezclas de cloroformo-metanol, en una proporción de 2 a 1, y luego disolvía los gangliósidos en agua. Esos métodos clásicos se han complementado con las técnicas de cromatografía. Por su sencillez y rapidez, la cromatografía en capa fina, mono- o bidimensional, usando diferentes líquidos de desarrollo, es de gran utilidad. No obstante, la identificación de un gangliósido solamente por esta técnica se considera insuficiente, sobre todo para tejidos extraneurales, dada la escasa concentración de gangliósidos en ellos y la posible coincidencia en el desplazamiento de varios gangliósidos. Asimismo, la separación de distintos gangliósidos, dada su mayor o menor polaridad, se ha logrado mediante cromatografía con derivados de cambio iónico débiles.

Para la caracterización de un gangliósido se acude a técnicas que por su especificidad y sensibilidad permiten identificar la naturaleza o la determinación cuantitativa de los constituyentes: ácidos grasos, esfingoide y restos glucídicos; además de establecer, generalmente, la posición del enlace y la forma anomérica. Las hidrólisis secuenciales mediante reactivos adecuados y, especialmente, las llevadas a cabo con la participación de hidrolasas específicas han resultado útiles. Las técnicas de permetilación, reducción con tetrahidruro de boro y sodio se emplean también con este fin. Desde hace algún tiempo, se aprovecha el desarrollo de otras técnicas: cromatografía de alta resolución, espectrometría de masas y resonancia magnética nuclear.

En cuanto a ensayos cuantitativos, los métodos espectrofotométricos para determinar los constituyentes glucídicos, y de modo especial la concentración en ácidos siálicos por ser un componente característico de los gangliósidos, resultan imprescindibles. Mas, a pesar de existir varios métodos para valorar este compuesto, la determinación exacta del mismo no es tarea fácil, dadas las interferencias ocasionadas por otros glúcidos, pérdidas durante las hidrólisis y otros obstáculos. Recientemente se han introducido las técnicas destinadas a la preparación de anti-

ALGUNOS ANTECEDENTES HISTORICOS SOBRE GANGLIOSIDOS		
AÑO	INVESTIGADOR	INVESTIGACION O RESULTADO MAS DESTACADO
1935-42	KLENK	Aislamiento de gangliósidos a partir de cerebros inicialmente de pacientes de las enfermedades de Niemann-Pick y Tay-Sachs.
1936-50	BLIX y col.	Hallazgo de una hexosamina como constituyente de los gangliósidos cerebrales.
1951	YAMAKAWA y SUZUKI	Aislamiento de un gangliósido (hematósido) a partir de eritrocitos de caballo.
1951	FOLCH y col.	Aislamiento de la "estrandina".
1958	BRADY y col.	Participación de derivados nucleotídicos en la biosíntesis de gangliósidos.
1959	BOGOG	Interacciones entre virus de la gripe y los gangliósidos cerebrales.
1960-62	KLENK y col.	Determinación del peso molecular de varios gangliósidos y de la estructura del hematósido.
1962 1963	DAIN y col.; SVENNERHOLM; KOREY	Separación de gangliósidos por cromatografía en capa fina.
1963-72	GATT y col.	Degradación de gangliósidos en el cerebro.
	KLENK y col.; KUHN y WIEGANDT	Establecimiento de la estructura definitiva de varios gangliósidos cerebrales.
1964	SWEELEY y WALKER	Determinación de glúcidos en gangliósidos por cromatografía en fase gaseosa.
1964	HAKOMORI	Determinación de las posiciones de los enlaces glicosídicos por permetilación, y otras investigaciones relacionadas.
1966	CAPUTTO y col.	Intervención de una sialiltransferasa en la biosíntesis de gangliósidos; disposición de éstos en membranas neuronales.
1970	ROSEMAN	Rutas de biosíntesis de gangliósidos con intervención del "complejo de la multiglicosiltransferasa".
1970	SVENSSON	Determinación de las posiciones de los enlaces glicosídicos por cromatografía en fase gaseosa y espectrometría de masas.
1970	LEDEEN, YU y ANDO	Caracterización de gangliósidos extraneurales y neurales, y revisión de datos sobre gangliósidos.
1971	BRADY	Sistematización de deficiencias enzimáticas que ocasionan acumulación de gangliósidos y cerebrósidos; terapia por reemplazamiento.
1973	LI y LI	Utilización de varias glicosidasas para determinar la secuencia y tipo de enlace de la fracción oligosacarídica de los gangliósidos.
1973	TETTAMANTI y col.	Metabolismo, función y aplicaciones de gangliósidos cerebrales.
1975	RAHMANN y col.	Funciones y filogenia de los gangliósidos, y estudios comparativos sobre su distribución.
1976	DREYFUS y col.	Patrón de gangliósidos de la retina en función de la edad.
1978	HARRIS	Aplicación de las técnicas de resonancia magnética nuclear al conocimiento de las estructuras de gangliósidos.
1978-80	IWAMORI; SVENNERHOLM y col.; MANDEL y col.	Resolución de mezclas de gangliósidos por técnicas perfeccionadas de cromatografía de cambio iónico y capa fina.
1980	RAPPORT y col.; MARCUS y col.	Anticuerpos frente a gangliósidos.
1980	SVENNERHOLM y col.; HAYWOOD y col.; TOMASI y col.	Función de los gangliósidos en la transmisión sináptica y como receptores de virus Sendai o toxinas (colérica, tetánica).
1981-83	GALJAARD y col.; ROBINSON (D.)	Factor corrector de la deficiencia combinada de β-galactosidasa/neuraminidasa y estudios sobre proteína protectora de la proteolisis.
1981-83	GORIO y col.	Los gangliósidos como agentes para el tratamiento de neuropatía diabética y como factores de la diferenciación, crecimiento y regeneración neuronal.
1982	KOPROWSKI y col.; GINSBURG	Producción de anticuerpos monoclonales correspondientes a fucogangliósidos asociados a cáncer gastrointestinal.

3. LOS GANGLIOSIDOS O SIALOGLICOESFINGOLIPIDOS fueron descubiertos por E. Klenk, de la Universidad de Colonia, hace unos cincuenta años, quien pensó inicialmente que se trataba de un único gangliósido. Por haberlo extraído de células *gang*lionares (neuronas grandes) y contener moléculas glucídicas (*osas*), lo denominó "Gangliosid". Hacia 1970 se conocía ya la estructura de una veintena de ellos, la investigación no se ha cerrado.

cuerpos monoclonales frente a monosialogangliósidos, entre otros motivos por el interés que presentan algunos fucogangliósidos como antígenos hallados en cáncer gastrointestinal.

Patrón y concentración

Se conoce por "patrón de gangliósidos" el conjunto de los porcentajes de gangliósidos determinados cualitativa y cuantitativamente en un material. Puede referirse a los que se hallan en mayor proporción. Estos gangliósidos mayoritarios, para órganos neurales de mamíferos, son el GM_1, GD_{1a}, GD_{1b} y GT_{1b}, que pueden alcanzar del 65 por ciento al 85 por ciento del contenido total del cerebro de mamíferos. O puede referirse dicho patrón conjuntamente a los porcentajes de gangliósidos mayoritarios y minoritarios.

El patrón de gangliósidos depende de varios factores. En primer lugar, de la especie animal y la edad del individuo. También depende del órgano o tejido analizado. A este respecto, trabajos efectuados por P. Hueso, en colaboración con A. Reglero y uno de los autores (Cabezas), con linfocitos procedentes de bazo y de sangre periférica de cerdo, muestran patrones de gangliósidos muy diferentes entre sí. Incluso para un mismo órgano o tejido se han encontrado diferencias según la fracción subcelular analizada. Por último, el patrón queda condicionado a circunstancias diversas, patológicas o no. En ensayos realizados en nuestro departamento de bioquímica de la Universidad de Salamanca, M. Rodrigo ha deducido que los patrones de gangliósidos de cerebro, cerebelo y bulbo de ratas sometidas a ingestión de etanol durante varios meses, o de ratas a las que se ha inyectado el analgésico fuerte pentazocina, experimentan algunas modificaciones de los respectivos patrones determinados en ratas no sometidas a estos tratamientos. El patrón para un material dado es el reflejo de los mecanismos de biosíntesis y de degradación de gangliósidos allí confluyentes.

El patrón de gangliósidos de las especies de mamíferos presenta, para los cuatro gangliósidos mayoritarios y en condiciones normales, valores similares, lo que podría explicarse como motivado por el desempeño de unas mismas funciones. Por el contrario, el patrón de gangliósidos de peces, anfibios y reptiles es más heterogéneo, debido quizás a que el mismo refleja no sólo aspectos filogenéticos de organización del sistema nervioso, sino también su participación en funciones de adaptación a distintas temperaturas. La existencia de gangliósidos con numerosos restos de ácidos siálicos, estudiados en peces, y de gangliósidos de estructura inhabitual son ejemplos que confirman

esta diversidad. El patrón de gangliósidos puede modificarse independientemente de la concentración de los mismos; ésta suele expresarse en microgramos o en nanomoles de ácido siálico por gramo de peso de tejido.

Biosíntesis

Brady y colaboradores dieron, en 1958, un gran paso adelante en el descubrimiento de las rutas metabólicas de biosíntesis de los gangliósidos con su trabajo sobre la transferencia de un resto de galactosa desde el derivado nucleotídico uridin-difosfato-galactosa, UDP-gal, a un aceptor; este proceso tenía lugar en la fracción microsómica de rata joven. Suponía esto la confirmación de la existencia de unos mecanismos comunes con los de la biosíntesis de oligosacáridos.

Desde comienzos de la década de 1970, varios equipos han tratado de esclarecer el mecanismo o mecanismos referentes a la biosíntesis de tales compuestos. El grupo norteamericano de S. Roseman, quien había establecido ya la biosíntesis del ácido *N*-acetil-neuramínico (NeuAc), al igual que lo había hecho L. Warren, logró inicialmente importantes resultados. En Argentina, R. Caputto y sus colaboradores han establecido las rutas de biosíntesis de los gangliósidos hoy admitidas y la localización subcelular de tales procesos. En Italia, G. Tettamanti y colaboradores han dedicado su atención también a este tema. Complementarios de estos trabajos son otros sobre aislamiento y caracterización de diversos gangliósidos, con los que han destacado R. W. Ledeen y R. K. Yu.

Este conocimiento incompleto es debido a la enorme diversidad y complejidad estructural de los mismos, la dificultad y limitación de las técnicas, la existencia de varias vías hipotéticamente válidas para explicar la biosíntesis de un mismo gangliósido y, por último, al error en que cayeron inicialmente diferentes equipos de investigación al considerar que los gangliósidos se biosintetizarían donde se localizan.

Expondremos las líneas generales de la síntesis metabólica de los gangliósidos. La porción glucídica de los sialoglicoesfingolípidos se forma en el aparato de Golgi. Algunos sostienen la posibilidad de que ocurra también en el retículo endoplásmico y en la membrana citoplasmática. A partir de la ceramida, la incorporación de restos glucídicos tiene lugar, de una manera secuencial, mediante cesión de éstos por los respectivos uridindifosfato-hexosa y hexosamina (UDP-hexosa y UDP-hexososamina), citidinmonofosfato-ácido siálico (CMP-ácido siálico) y, excepcionalmente, guanosindifosfato-L-fucosa (GDP-fucosa); estas reacciones las catalizan unas 15 o 20 glicosiltransferasas, enzimas integradas en lo que Roseman ha denominado "complejo de la multiglicosiltransferasa".

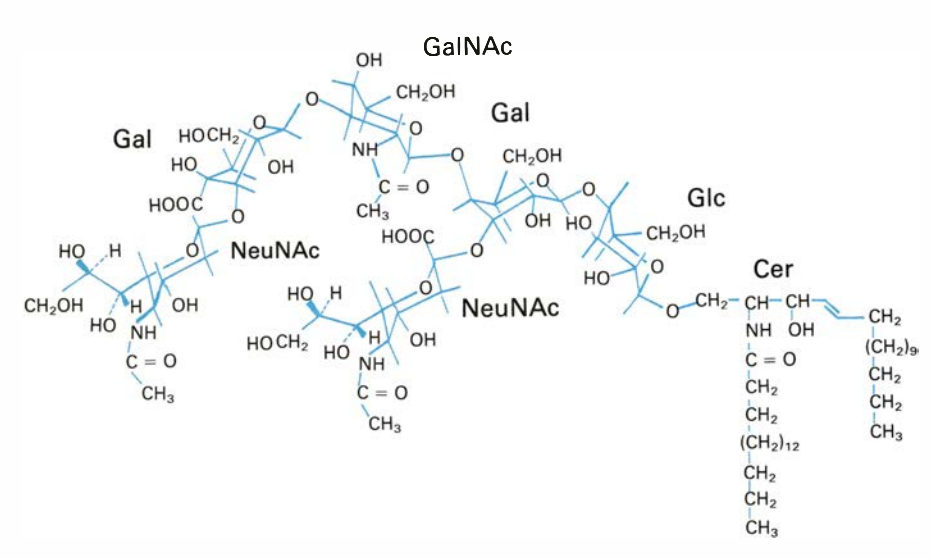
En este proceso, la doble especificidad de cada enzima respecto al monosacárido-nucleótido y al aceptor garantiza una gran constancia en la composición de los gangliósidos de un material dado; no obstante, pueden biosintetizarse gangliósidos más o menos complejos. Una confirmación indirecta de la intervención de mecanismos de especificidad de especie –o sea, característicos de cada especie– en la biosíntesis de gangliósidos ha sido apreciada en nuestro departamento por García-Alonso, Reglero y uno de los autores (Cabezas) al obtener el patrón de gangliósidos de distintas partes de encéfalo de caballo, asno y mulo, que resultó ser generalmente intermedio para el híbrido mular respecto a los de caballo y asno. Los principales pasos de la ruta metabólica de biosíntesis de gangliósidos se recogen en la figura 8.

El uso de aceptores exógenos o de aceptores endógenos en los ensayos experimentales ha conducido a algunas diferencias parciales del esquema aquí esbozado: gangliósido GM_1 como precursor del GD_{1a} o, conjuntamente, del GD_{1a} y del GD_{1b}. Por otro lado, la biosíntesis del gangliósido GT_{1c} no parece poderse explicar a partir del GD_{1b} (ni menos, a partir de GD_{1a}), por lo que Yu y Ando han sugerido la existencia de dos rutas confirmadas y una tercera como muy probable.

También Rahmann e Hilbig han aceptado y ampliado recientemente el esquema de Yu y Ando en relación con sus estudios filogenéticos. Comprendería tres vías: *a, b* y *c*; de las que la *a* sería utilizada por los mamíferos más evolucionados; las *b* y *a*, por reptiles y aves; las *b* y *c*, por anfibios y peces cartilaginosos; y la *c*, por teleósteos.

Recientes trabajos de Fishman y colaboradores indican que la glicosilación secuencial de un gangliósido para originar otro más complejo se realiza con la participación de un pequeño acervo de intermediarios. Estos resultados guardan analogía con los de Caputto y colaboradores, según los cuales existirían dos pequeños acervos en el cerebro, siendo efectuada la biosíntesis de cada gangliósido por su sistema multienzimático, con la intervención de un acervo limitado de intermediarios de transición, que garantizarían la separación, de los formados, de los de igual estructura preexistentes; los primeros pasarían a otras posiciones de la membrana, de las que no podrían volver para convertirse en sustratos de ulteriores biosíntesis. En este sentido, los resultados de Fishman y colaboradores ponen de manifiesto que la transferencia de gangliósidos desde el sitio de su biosíntesis hasta la superficie celular requeriría unos 20 minutos, siendo transportados a un ritmo constante hasta la membrana plasmática. En el caso de las células nerviosas, los gangliósidos sintetizados en el aparato de Golgi son transportados a lo largo del axón hasta las terminaciones nerviosas.

La biosíntesis de gangliósidos en circunstancias patológicas muestra que una deficiencia en la UDP-*N*-acetilgalactosaminil: GM_3-transferasa ocasiona un acúmulo anormal del gangliósido GM_3, fenómeno similar al de la simplificación de gangliósidos que sucede en las membranas de células eucarióticas después de la transformación



4. ESTRUCTURA DE UN GANGLIOSIDO: GD_{1a}. Se ha supuesto que el GM_1 sería su precursor. Nótese aquí la presencia de dos restos de ácido siálico no contiguos y de la cadena larga de osas-osamina.

de éstas por virus oncogénicos, tratamientos con agentes carcinógenos, etcétera. La explicación lógica de estos hechos radica en el papel crucial de esta enzima en la ruta de biosíntesis que conduce a otros gangliósidos más complejos. Según Dawson y colaboradores, la biosíntesis de gangliósidos puede ser inhibida por β-endorfina, encefalinas y opiáceos.

Degradación

Bastante mejor, y desde hace más tiempo, se conocen los procesos de degradación de gangliósidos. Lo que en parte se explica porque se entendía mejor la naturaleza de las hidrolasas que la de las glicosiltransferasas. Nuevos aspectos surgidos últimamente, así el de la intervención de factores proteicos, han contribuido a afianzar la idea de que todavía restan importantes áreas por esclarecer. El acúmulo anormal de ciertos esfingolípidos pudo relacionarse, ya en los primeros años de estas investigaciones, con deficiencias en algunas enzimas o isoenzimas implicadas en su catabolismo, al tiempo que se incrementa la actividad de otras glicosidasas o de alguna isoenzima.

Como Gatt estableció en 1963, la eliminación del resto terminal de ácido siálico es la primera etapa de la degradación de los polisialogangliósidos. Secuencialmente van degradándose los esfingolípidos resultantes merced a la actividad catalítica de las restantes glicohidrolasas [*véase la figura 9*].

En el sistema nervioso, los gangliósidos con más de un resto de ácido siálico son degradados hasta GM_1 en las terminaciones nerviosas, siendo éste transportado a lo largo del axón hasta los lisosomas, donde resulta degradado.

Comprenderemos mejor el catabolismo o degradación de los gangliósidos hasta otros metabolitos más sencillos si resumimos algunas investigaciones recientes al respecto. Carubelli y colaboradores han aislado y caracterizado diversos tipos de neuraminidasa o sialidasa de cerebro que presentan distinta localización subcelular y poseen propiedades muy diferentes. Ello significa que ciertos compuestos pueden ser sustratos de unas neuramidasas y no de otras. La degradación artificial de gangliósidos como el GD_{1a} mediante neuraminidasa de *Vibrio cholerae* empieza por la liberación del componente ácido siálico externo, originándose así el gangliósido GM_1. Este producto resultante es insensible a la actividad de dicha enzima. Solamente después de liberarse la *N*-acetil-β-D-galactosamina por la

TERMINOLOGIA DE GANGLIOSIDOS	
ESTRUCTURA	SVENNERHOLM, 1964
CERAMIDA — glc — gal — NANA	GM_3
CERAMIDA — glc — gal — gal. Nac NANA	GM_2
CERAMIDA — glc — gal — gal. Nac — gal NANA	GM_1
CERAMIDA — glc — gal — gal. Nac — gal — NANA NANA	GD_{1a}
CERAMIDA — glc — gal — gal. Nac — gal NANA — NANA	GD_{1b}
CERAMIDA — glc — gal — gal. Nac — gal — NANA NANA — NANA	GT_1

Abreviaturas: NANA = ácido N-acetilneuramínico; glc = glucosa; gal = galactosa; gal. Nac = N-acetilgalactosamina. (Desde hace pocos años, se recomienda usar las abreviaturas NeuAC o NeuNAc en vez de NANA, para el ácido N-acetilneuramínico).

5. NOMENCLATURA DE SVENNERHOLM de los gangliósidos que, entre las distintas propuestas, ha acabado por imponerse. La letra *G* designa el término gangliósido; *M*, *D* y *T* abrevian el prefijo griego de los restos de ácido siálico; a saber: monosialogangliósidos, disialogangliósidos y trisialogangliósidos.

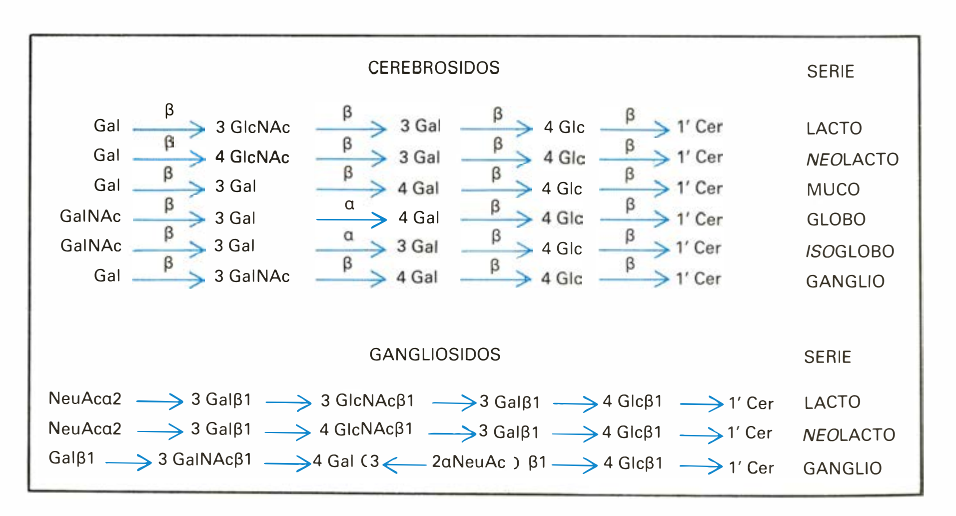
hexosaminidasa correspondiente puede actuar la neuraminidasa de *V. cholerae* (y la de casi todos los virus que poseen dicha actividad) e hidrolizar el derivado del GM_1. No obstante, la neuraminidasa procedente de *Clostridium perfringens* es capaz de actuar directamente sobre el GM_1 e hidrolizarlo.

Otros datos sobre degradación

La localización principal de las enzimas que participan en los procesos degradativos es lisosómica. Quedan todavía por averiguar bastantes aspectos referentes a los lisosomas y a las enzimas lisosómicas. A veces su complejidad de acción es muy elevada. Por ejemplo, la hexosaminidasa A, cuya carencia origina la enfermedad de Tay-Sachs por acúmulo de su sustrato, el gangliósido GM_2, está compuesta por dos tipos de subunidades diferentes, α y β; como indican Srivastava y colaboradores, los sitios activos estarían en las subunidades β, mientras que las subunidades α funcionarían como modificadoras de la especificidad, dirigiendo la enzima hacia el gangliósido GM_2. En esta enfermedad, la carencia genética sólo afecta a las subunidades α, ya que las β se reagrupan entre sí para incrementar los niveles de hexosaminidasa B.

Artificialmente se utilizan detergentes para favorecer la hidrólisis de gangliósidos; *in vivo* se produce esta degradación participando factores de activación, de naturaleza proteica, con tamaño molecular próximo a 20.000 y localización lisosómica, para favorecer la actividad de la neuraminidasa, β-*N*-acetilhexosaminidasa, β-D-galactosidasa y otras enzimas relacionadas.

Una cuestión no menor es la de si pueden degradar a las glicoproteínas las mismas glicosidasas que hidrolizan los gangliósidos. La respuesta es que se conocen glicosidasas capaces de actuar



6. SERIES O FAMILIAS a las que pertenecen glicoesfingolípidos neutros o cerebrósidos (*arriba*) y ácidos gangliósidos (*abajo*). El gangliósido que posea galactosa en vez de glucosa se adscribe a la serie GALA.

NOMENCLATURAS ACTUALES DE ALGUNOS GANGLIOSIDOS			
Sven.	NOMENCLATURA IUPAC-IUB	Sven.	NOMENCLATURA IUPAC-IUB
GM_4	I^3 αNeuAc-GalCer	GD_{1a}	II^3 αNeuAcIV^3 αNeuAc-GgOsa$_4$ Cer
GM_4	I^3 αNeuAc-GlcCer	GD_{1b}	II^3 α(NeuAc)$_2$-GgOsa$_4$ Cer
GM_3	II^3 αNeuAc-LacCer	LD_1	IV^3 (NeuAc)$_2$-nLcOsa$_4$ Cer
GM_3	II^3 αNeuGc-LacCer	GT_{1a}	IV^3 (NeuAc)$_2$ II^3 NeuAc-GgOsa$_4$ Cer
GD_3	II^3 α(NeuAc)$_2$-LacCer	GT_{1b}	IV^3 NeuAc, II^3 (NeuAc)$_2$-GgOsa$_4$ Cer
GD_3	II^3 α(NeuGc)$_2$-LacCer	GT_{1c}	II^3 (NeuAc)$_3$-GgOsa$_4$ Cer
GD_3	II^3 α(NeuAc, NeuGc)-LacCer	GQ_{1b}	IV^3 (NeuAc)$_2$, II^3 (NeuAc)$_2$-GgOsa$_4$ Cer
GM_2	II^3 αNeuAc-GgOsa$_3$ Cer	GQ_{1c}	IV^3 NeuAc II^3 (NeuAc)$_2$-GgOsa$_4$ Cer
GD_2	II^3 α(NeuAc)$_2$-GgOsa$_3$ Cer	GP_{1b}	IV^3 (NeuAc)$_3$ II^3 (NeuAc)$_2$-GgOsa$_4$ Cer
GM_1	II^3 αNeuAc-GgOsa$_4$ Cer	GP_{1c}	IV^3 (NeuAc)$_2$ II^3 (NeuAc)$_3$-GgOsa$_4$ Cer
LM_1	IV^3 αNeuAc-nLcOsa$_4$ Cer	—	VI^3 αNeuAc-nLcOsa$_6$ Cer
—	IV^3 αNeuGc-nLcOsa$_4$ Cer		
ABREVIATURAS: Sven. = Svennerholm. IUPAC-IUB = "International Union of Pure and Applied Chemistry-International Union of Biochemistry".			

7. CON EL FIN DE EXPRESAR adecuadamente la estructura de los lípidos, las recomendaciones de 1976 de la IUPAC-IUB (International Union of Pure and Applied Chemistry-International Union of Biochemistry) introdujeron una terminología que, no obstante su precisión, resultaba difícil de retener. De ahí que los bioquímicos sigan prefiriendo ordinariamente la de Svennerholm (*columna de la izquierda*).

sobre ambos grandes grupos de glicoconjugados, mientras que otras son exclusivas de cada uno de ellos; y, aun para cada grupo, pueden presentarse en ellas limitaciones y diferencias muy marcadas. Como era previsible, estas glicosidasas muestran una amplia variedad de posibilidades de actuación.

Burton y colaboradores observaron en 1964 que los gangliósidos eran de los compuestos más rápidamente metabolizados en el cerebro; calculó su vida media, de unos 24 días cuando se sintetizaban utilizando glucosamina, y, de unos 10 días, si el precursor era la galactosa. Si se inyectan gangliósidos marcados, desaparecen de la circulación más o menos rápidamente, según su naturaleza; el GD_{1a} tiene una vida media de unas cuatro horas, y el GM_1, de 21 horas.

Si durante los últimos siete lustros se ha venido concediendo a las proteínas y a los ácidos nucleicos un papel decisivo en la expresión funcional de los seres vivos y se ha considerado a los glúcidos como un material con misiones eminentemente energéticas o de sostén, desde la década del 70 se ha ido abriendo camino la idea de que ciertos glúcidos a lo menos también tienen funciones de significado biológico no inferior al de los otros grupos mencionados, si están integrados en glicoconjugados.

Función de los gangliósidos

Quizá la primera declaración precisa sobre la importancia de algunos glúcidos en estos cometidos biológicos la hiciera N. Sharon. Dice: "El comportamiento social de la célula viva –que comprende la comunicación intercelular, la regulación del crecimiento y de la diferenciación, la respuesta inmune y quizá también la aparición de tumores malignos– está mediado principalmente por la superficie de la célula y, más concretamente, por las moléculas de azúcar ramificadas, de las que está tachonada toda la superficie celular" ["Lectinas", INVESTIGACIÓN Y CIENCIA, febrero de 1977]. Y en otro lugar señala que la especificidad de muchos polímeros naturales está escrita en términos de residuos glucídicos.

Gran parte de las funciones de los gangliósidos están relacionadas con su participación en la transmisión del impulso nervioso. Sus efectos estimuladores de la misma estarían causados por su capacidad de fijación de Ca^{++} en sus restos de ácidos siálicos. Como es sabido, el Ca^{++} es la señal que va a permitir la fusión de las vesículas sinápticas con la membrana presináptica y el posterior vertido, a la hendidura sináptica, de los neurotransmisores que contienen. Los gangliósidos están anclados en la membrana presináptica, con la porción glucídica orientada hacia el exterior. Así, podría fijarse el Ca^{++} externo sobre los ácidos siálicos, como paso previo a su internalización por el canal correspondiente, para inducir la fusión de las vesículas.

Los polisialogangliósidos, que son los gangliósidos más abundantes en las vesículas sinápticas, podrían participar también en la facilitación de la fusión de estas vesículas con la membrana presináptica, debido a sus características biofísicas de interacción lipídica.

Según Rahmann, durante el estado de potencial de reposo, en un terminal nervioso la membrana presináptica está cerrada, teniendo complejos lipofílicos gangliósidos-Ca^{++} en la superficie externa. Cuando llega un impulso nervioso, el potencial eléctrico de la membrana va cambiando según avanza la onda de despolarización, por movimiento de cationes (Na^+ y K^+). Los iones K^+ que alcanzan la parte externa de la membrana presináptica tienden a disociar, de los gangliósidos, el Ca^{++}. Así, estos gangliósidos se hacen más hidrofílicos, originando un incremento de la permeabilidad de la membrana. Esta apertura parcial de la membrana presináptica posibilita una liberación de neurotransmisores a la hendidura sináptica. Por encima de una concentración crítica, el transmisor liberado, acetilcolina, disocia más Ca^{++} de los complejos con los gangliósidos, de un modo cooperativo con el K^+. Este incremento rápido de la permeabilidad local origina la liberación total de los neurotransmisores disponibles, que continuarán su interacción en la membrana postsináptica. Paralelamente a la repolarización posterior de las membranas pre y postsinápticas, se forman de nuevo los complejos Ca^{++}-gangliósidos, para empezar el mecanismo cíclico, quedando entonces la membrana presináptica cerrada, con un carácter más lipofílico.

Los gangliósidos también podrían participar en la recaptura del K^+ (expulsado en la despolarización nerviosa), fijándolo transitoriamente en sus sitios acídicos (de sus ácidos siálicos); este efecto podría ser inhibido por proteínas básicas, que interferirían el mecanismo de fijación iónica.

Parece ser que están igualmente relacionados con la actividad de la ATP-asa que transporta K^+ y Na^+ a través de la membrana, ya que dicha enzima está rodeada de altas concentraciones de gangliósidos. También participarían en la recaptura de neurotransmisores que, una vez liberados a la hendidura sináptica, no se hubieran fijado directamente a su receptor postsináptico. La acetilcolina y otras aminas que funcionan como neurotransmisores están cargadas positivamente, y el retrotransporte de éstas estaría facilitado por las cargas negativas de los ácidos siálicos de los gangliósidos anclados en la membrana presináptica.

Memoria

Siguiendo con estos modelos, los gangliósidos estarían implicados en algunos de los procesos neurofisiológicos que integran la memoria. Para explicar ese fenómeno, se ha considerado que hay pequeñas regiones limitadas de sinapsis que se modificarían al objeto de facilitar la liberación de los neurotransmisores. Esta modificación podría estar

causada por un aumento en la cantidad de sialoglicoconjugados. La presencia de más complejos Ca^{++}-gangliósidos produciría una liberación mayor y más rápida del neurotransmisor.

Si los procesos indicados se repiten frecuentemente de la misma forma, como en una situación de aprendizaje, se produciría una modificación de esta región especial de contacto molecular, quizá como un cambio de configuración en la composición de micelas de gangliósidos. Esto podría ser la base para una facilitación de una vía neuronal, que podría ser el sustrato morfológico de un enagrama. El mantenimiento de las vías neuronales formadas (enagramas), por largos períodos de tiempo, lo garantizan los procesos eléctricos continuos. El recuerdo (o reconsideración del enagrama) podría explicarse por inhibición presináptica; la información eléctrica se enfocaría así sólo sobre esos sitios específicos de contacto sináptico que habrían sido facilitados por el aumento de sialoglicoconjugados.

Los gangliósidos también intervienen en la adaptación de especies que viven a bajas temperaturas. Así, se ha observado un aumento de la síntesis *de novo* de polisialogangliósidos durante dicha adaptación: cuanto más fría sea la temperatura ambiente, habrá mayor proporción de polisialogangliósidos.

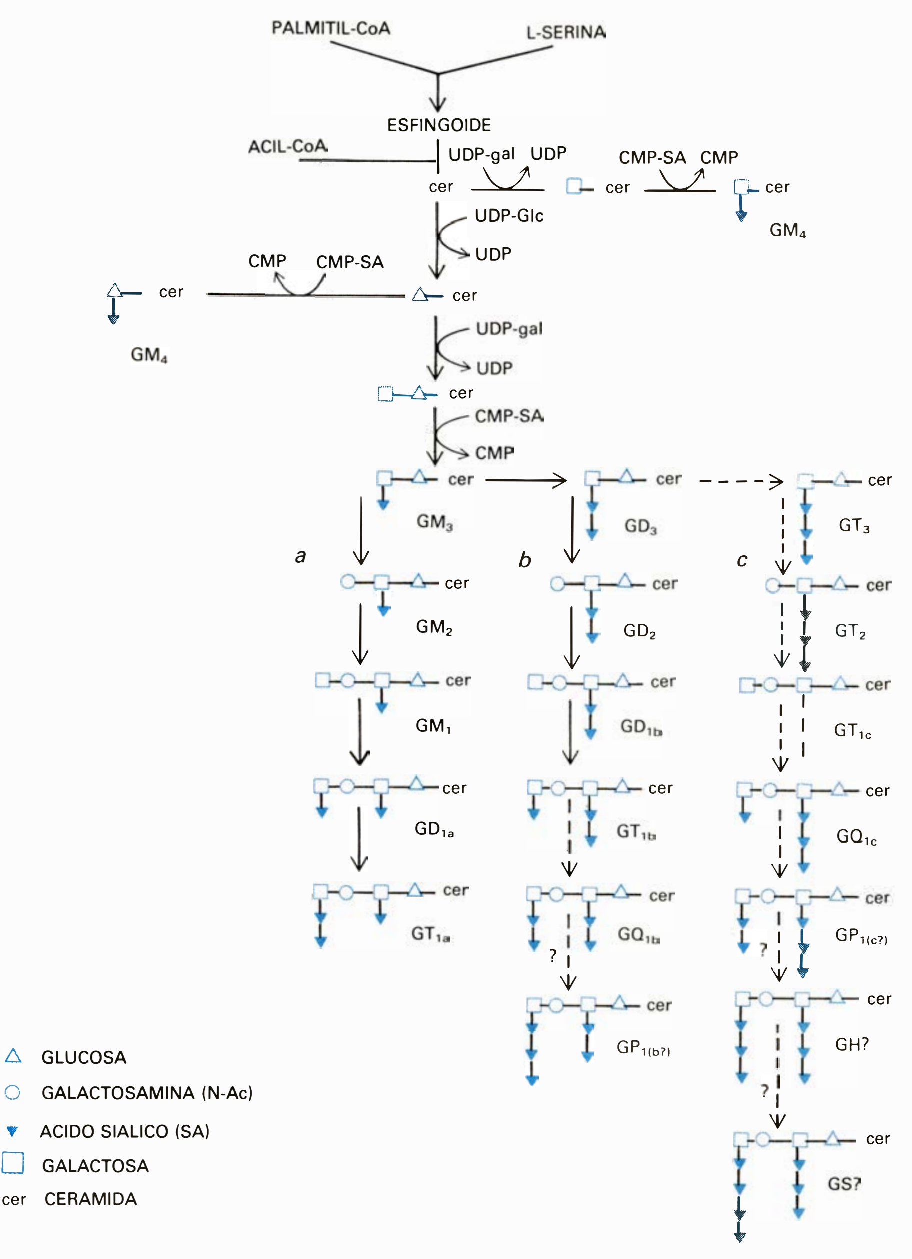
Con los complejos Ca^{++}-gangliósidos formados, la membrana presináptica estaría cerrada, pero su disociación haría que se abriera. Los complejos gangliósidos-Ca^{++} son tanto más estables cuanto más restos de ácidos siálicos contengan. Como los complejos lípidos-Ca^{++} son menos estables a bajas temperaturas, parece probable que la estabilidad disminuida de los complejos gangliósidos-Ca^{++} se compense con la formación de gangliósidos más polares, por la presencia de más residuos de ácidos siálicos en ellos. Al tener más grupos polares, negativos, por los ácidos siálicos adicionales, tendrán mayor capacidad de fijación de iones Ca^{++}, compensando así las bajas temperaturas y haciendo funcionales a las membranas sinápticas de nuevo.

Fijación de toxinas

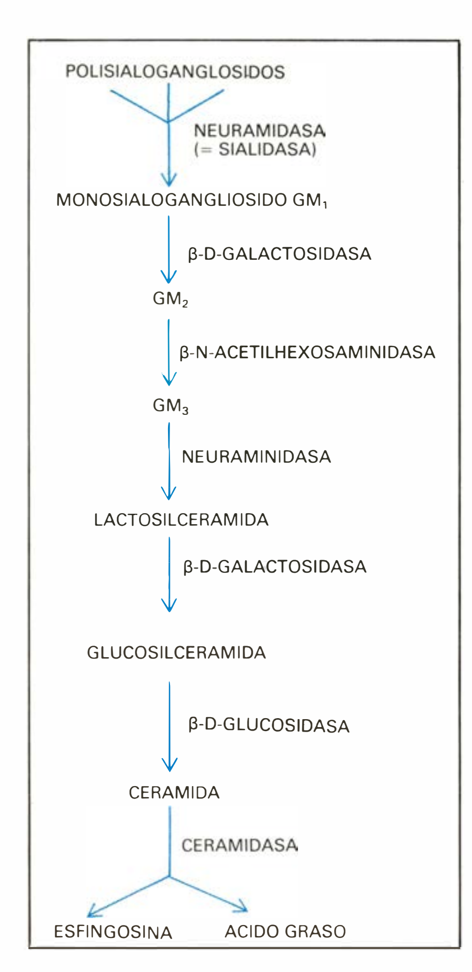
Los gangliósidos participan también en la fijación de toxinas. El estudio del mecanismo de acción de la toxina del cólera (colerágeno) y de su interacción con el gangliósido GM_1 es de gran importancia, no sólo por corresponder a la mayor parte de los cuadros diarreicos, sino también porque mimetiza el mecanismo fisiológico de interacción de receptores de membrana de naturaleza gangliosídica, con distintas hormonas de tipo glicoproteínico. El colerágeno ejerce su acción a través de una activación mantenida de la adenilciclasa en el intestino.

El suceso inicial es la fijación de las subunidades B de la toxina sobre el gangliósido GM_1, que actúa como receptor en la superficie celular, lo que posibilita que a continuación la subunidad A tenga acceso a la membrana, penetre y cataiice la ribosilación por ADP, dependiente de NAD, de restos de arginina de la subunidad reguladora de la adenilciclasa, caracterizada por fijar GTP para la activación del sistema. Esto se produciría en la superficie interna de la membrana celular, o en el interior. Al quedar modificado covalentemente por dicha ribosilación este componente regulador del complejo de la adenilciclasa, queda inhibida su capacidad GTP-ásica intrínseca; y, al no producirse la hidrólisis del GTP, éste deja al sistema de la adenilciclasa en un estado de activación permanente, sin que se produzca el retrocontrol de la reversión de la ciclasa de su forma activa a la inactiva. Como consecuencia de esto, el AMPc es sintetizado de forma continua por el sistema, sin que las fosfodiesterasas puedan inactivarle a tiempo, dada su producción masiva. Al aumentarse, por efecto del AMPc, el transporte del Na^+ de forma incontrolada y continua, este ion sale profusamente a la luz del intestino, arrastrando cantidades masivas de agua, que es lo que produce las graves diarreas.

El componente A_1 de la subunidad A es el que realmente está dotado de la capacidad enzimática de activación del



8. BIOSINTESIS DE GANGLIOSIDOS. Las flechas de trazo continuo señalan las rutas comprobadas, en tanto que las de trazo discreto, las no confirmadas, aunque muy probables; *a* sería la ruta utilizada por los mamíferos; *b* y *a*, por los reptiles y aves; *c*, por peces teleósteos. (Según postulan Rahman e Hilbig.)



9. DEGRADACION de gangliósidos (*esquema general*). Se conoce desde hace tiempo. Como Gatt estableció en 1963, la eliminación del resto terminal del ácido siálico constituye la primera etapa de la degradación de los polisialogangliósidos. Los esfingolípidos resultantes se van degradando después por la actividad catalítica de las glicohidrolasas.

sistema; quizá se libere del componente A_2, por reducción del puente disulfuro que los une, para ser activo. Posee actividad NAD-glicohidrolásica y ADP-ribosiltransferásica.

Se ha propuesto que algunos tipos de gangliósidos (GD_{1b}, GT_{1b}) funcionan como receptores de la toxina tetánica, estando implicados en una función que sería inhibida por la toxina. La toxina tetánica actuaría en la membrana presináptica obstaculizando la liberación del neurotransmisor en el terminal nervioso, probablemente al interaccionar con dichos gangliósidos, bloqueando así la transmisión del impulso. El Ca^{++} ejerce un efecto antagónico en la fijación de la toxina tetánica, lo que indicaría que la toxina y el Ca^{++} competirían por sitios relevantes fisiológicamente.

La toxina botulínica bloquea la transmisión neuromuscular a nivel presináptico e impide la liberación de neurotransmisor. Parece ser que el gangliósido GT_{1b} está implicado en la unión de la toxina a las membranas presinápticas, actuando como receptor, o formando parte de su estructura.

Muchas de las hormonas glicoproteínicas se unen a receptores específicos en la superficie celular y activan la adenilciclasa. Parece probable que la toxina del cólera puede estar subvirtiendo un mecanismo normal de transferencia de información a través de membranas celulares. La hormona estimuladora del tiroides está compuesta de dos subunidades, α y β. La subunidad α es común en muchas hormonas glicoproteínicas, mientras que la subunidad β confiere la especificidad para el órgano diana y contiene, probablemente, los determinantes primarios para la fijación a las membranas. Esto guarda analogías con el mecanismo de acción de la toxina del cólera. Hay regiones con homología en la secuencia de aminoácidos del componente B de la toxina del cólera y de las cadenas β de las hormonas TSH, luteinizante, gonadotropina coriónica y FSH. Además, hay homología de secuencia entre la subunidad A_1 de la toxina del cólera y las cadenas α de estas hormonas. En contraste con el mecanismo de acción de la toxina, no se requiere NAD para la estimulación de la adenilciclasa por la TSH.

Después de que la hormona se hubiera unido al receptor glicopeptídico, interaccionaría con un gangliósido específico en la membrana y sufriría un cambio conformacional. El cambio en la estructura de la hormona promovería la estimulación de la adenilciclasa por la subunidad α. Pueden observarse de nuevo analogías entre este mecanismo de acción y el de la toxina del cólera. Las distintas hormonas indicadas reconocen secuencias oligosacarídicas distintas en las moléculas de gangliósidos, receptoras. Cada órgano diana tendría los receptores con la secuencia glucídica específica para la hormona a fijar.

Funciones inhibidoras

Varios gangliósidos inhiben la hemaglutinación vírica. Por tal se entiende la aglutinación que experimentan las hematíes en presencia de virus como el de la gripe. Se cree que tales gangliósidos proporcionarían un sitio receptor para los virus, que competiría con el sitio receptor de la superficie celular de naturaleza gangliosídica. Son necesarios receptores que contengan ácidos siálicos para que se produzca infección por algunos virus. Los gangliósidos inducirían la fusión célula-virus y su posterior penetración, después de fijarles en la superficie celular.

La acción antivírica del interferón de fibroblastos humanos y de ratón es inhibida por gangliósidos, que inhiben también la aglutinación de los eritrocitos RhO causada por los antisueros específicos. En los linfocitos, las cadenas oligosacarídicas de los gangliósidos son potentes determinantes antigénicos, y pueden ayudar a transmitir señales inmunológicas entre las células. Los linfocitos T podrían ejercer un efecto regulador sobre los linfocitos B, mediante la transferencia de gangliósidos. Estos intervendrían asimismo en la interacción linfocito-macrófago, que permite su cooperación.

Los gangliósidos están implicados en la regulación del crecimiento celular y los fenómenos tumorales. Una síntesis incompleta de gangliósidos superiores podría muy bien estar relacionada con una pérdida de la regulación del crecimiento demostrada en células transformadas. La síntesis de un glicolípido particular aumenta al establecerse un contacto célula-célula: la llamada "respuesta glicolipídica"; se la considera relacionada con la inhibición por contacto del crecimiento celular.

La respuesta glicolipídica y el proceso de inhibición por contacto están basados en una posible interacción entre proteínas similares a los anticuerpos y glicolípidos sensibles al contacto. Cuando dos células confluyen, esas proteínas y esos glicolípidos se enlazarían a través de estructuras complementarias. En las células malignas transformadas, las cadenas glucídicas estarían incompletas; no se formarían, pues, estructuras complementarias.

Las cadenas glucídicas de glicolípidos podrían actuar de receptores de un factor sérico de crecimiento. La inhibición del crecimiento celular inducida por anticuerpos monovalentes antigangliósidos podría estar causada por el bloqueo de la recepción del factor sérico de crecimiento. La inhibición del crecimiento en el contacto celular podría estar inducida por el encubrimiento del glicolípido receptor mediante una proteína complementaria.

El patrón de gangliósidos en tumores cerebrales difiere del correspondiente a tejido cerebral normal. En tumores se encuentran menos gangliósidos y se observa una disminución de la proporción relativa de polisialogangliósidos. Los gangliósidos menos polares suelen incrementarse en tumores malignos.

Tumores

Muchas líneas celulares que sufren transformación debida a virus oncogénicos muestran una simplificación en su patrón de gangliósidos. Los niveles de alguno de los más complejos (GD_{1a}) disminuyen, mientras que aumenta el del GM_3. Una *N*-acetilgalactosaminil

transferasa requerida para la biosíntesis de los gangliósidos complejos estaría ausente, o a muy bajo nivel. La actividad sialiltransferásica para la formación de disialogangliósidos está disminuida en tumores inducidos en rata. Los gangliósidos estarían implicados en los cambios de la superficie celular que se dan en la génesis tumoral.

Muchos anticuerpos monoclonales producidos por hibridomas obtenidos de ratones inmunizados con una línea celular de adenocarcinoma de colon humano tienen especificidad por tumores gastrointestinales humanos, fijándose a dichas células cancerosas. El antígeno para estos anticuerpos en la línea celular usada para la inmunización es un monosialogangliósido con fucosa, que ha sido aislado muy recientemente. Este gangliósido se ha detectado en adenocarcinomas gástricos y de colon, en carcinomas pancreáticos y en el meconio, pero no en tejidos adultos normales. La presencia de este gangliósido en el meconio sugiere que pudiera ser un antígeno embrionario asociado a tumores. Además, este antígeno se ha detectado en algunos tejidos embrionarios humanos, pero no en los tejidos adultos correspondientes.

Desarrollo

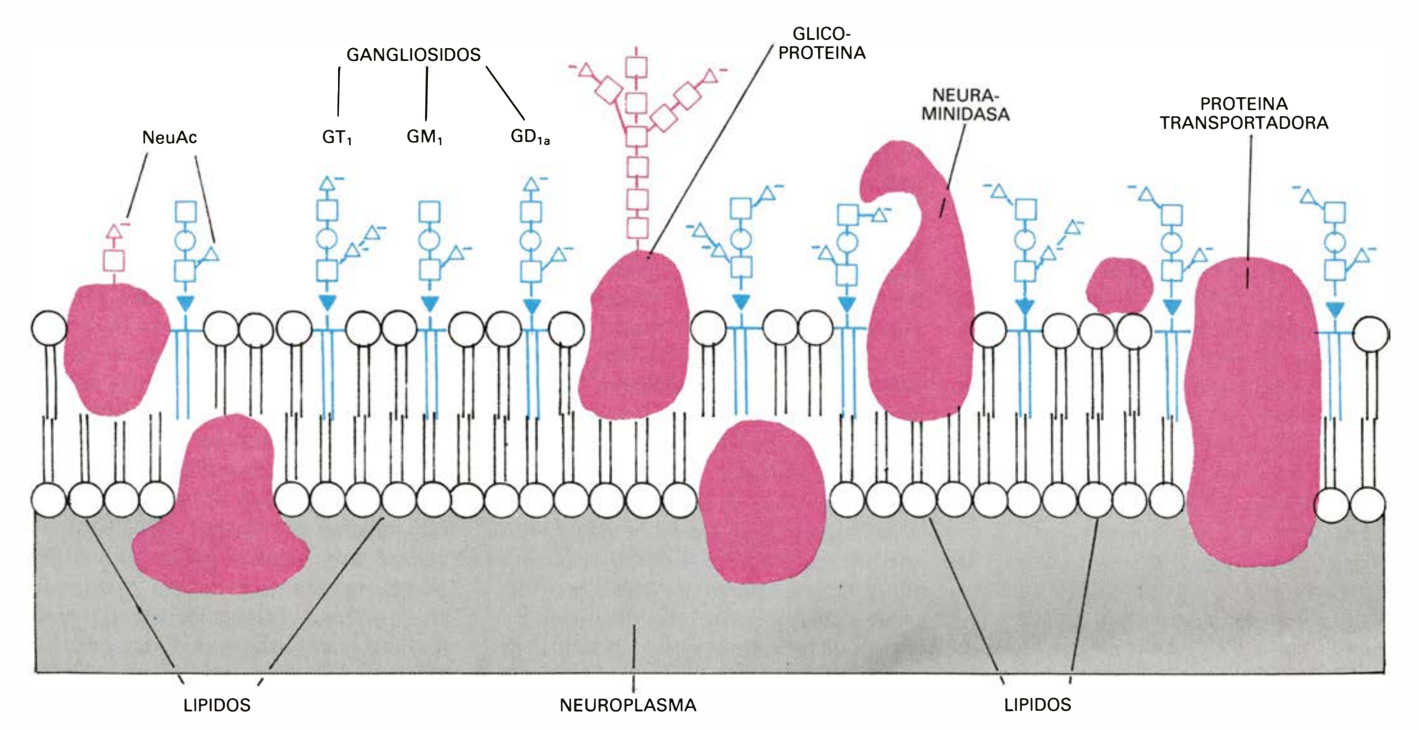
La composición de los gangliósidos cambia con el desarrollo. Vanier y colaboradores proponen tres grandes períodos. El primero se caracteriza por una multiplicación de neuronas y de células gliales y por un incremento moderado en la concentración de gangliósidos. Predominan GM_1 y GT_1, y empieza el incremento del GD_{1a}. Este tiempo se extiende hasta la semana 25 del feto humano. Durante el segundo período se asiste al crecimiento de dendritas y axones y al establecimiento de conexiones neuronales; corresponde al tiempo comprendido entre la semana 30 del feto humano y el nacimiento. Esta es la primera fase de rápido incremento de la concentración de gangliósidos. La tasa de crecimiento del GD_{1a} es máxima, mientras que la del GT_1 y GM_1 declina. La tercera fase tiene lugar en el hombre desde el nacimiento hasta los 8 meses. Se da una extensión ulterior de las conexiones neuronales y empieza la mielinización. En esta segunda ocasión de aumento rápido de gangliósidos se llega a estabilizar su patrón, siendo el GD_{1a} el predominante.

Entre los 4 y los 30 años de edad, disminuye el porcentaje de GD_{1a} y aumenta el de GD_{1b} y GT_1. La proporción relativa de GM_1, GD_{1a}, GD_{1b} y GT_1 no sufre ningún cambio significativo entre los 30 y los 70 años. Se considera que hay dos fases de acumulación rápida de gangliósidos: un período en el que se multiplican las neuronas y una segunda fase en la que se establecen conexiones intersinápticas. No obstante, el cambio en la concentración relativa de cada tipo de gangliósidos a lo largo de la vida depende del área del cerebro estudiada y de la especie.

La malnutrición durante fases del desarrollo temprano del individuo retarda la mielinización y el depósito de gangliósidos. La malnutrición incrementa la densidad de empaquetamiento celular y desorganiza la citoarquitectura de la corteza, efecto que es permanente. El pico normal de concentración de gangliósidos es menor y se retrasa varios días en esas condiciones. El retraso en la acumulación de gangliósidos parece estar relacionado con una disminución de la proliferación de fibras axodendríticas y con una maduración de la conducta retrasada.

El tratamiento con gangliósidos favorece en gran medida los procesos de regeneración y reinervación de fibras nerviosas. Al administrar gangliósidos se ha observado *in vitro* un aumento del alargamiento axonal. Los gangliósidos podrían funcionar como moléculas aceptoras de sustancias que promuevan el crecimiento en neuronas embrionarias. Después de su incorporación, podrían inducir modificaciones de membrana directa o indirectamente, lo que haría que un sitio determinado se hiciera susceptible para la formación de una nueva ramificación neuronal.

El tratamiento con gangliósidos ha demostrado ser efectivo igualmente en síndromes de trastornos mentales, mejorándose el rendimiento de los pacientes en tests neuropsicológicos. Con respecto a la diabetes, una de sus complicaciones más relevantes es la neuropatía que afecta a los sistemas motor, sensorial y autónomo. El tratamiento con gangliósidos permite la recuperación del tamaño axonal normal y potencia la velocidad máxima de conducción nerviosa.



10. POSICION DE ALGUNOS GANGLIOSIDOS, glicoproteínas y otros componentes en la membrana sináptica (a partir de Rahmann y col.). La concentración de lípidos en el sistema nervioso es bastante elevada. El funcionamiento adecuado del mismo depende especialmente de los esfingolípidos.

Juegos de ordenador

En el juego de la "guerra nuclear", dos programas hostiles entablan, sin ayuda externa, batallas de bits

A. K. Dewdney

Dos programas residentes en su hábitat nativo –los microcircuitos de memoria de un ordenador digital– se acechan mutuamente de una en otra posición de la memoria. Salen a veces de descubierta, en busca de su enemigo; otras, dejan caer barreras de bombas numéricas; en ocasiones se detienen para reparar los daños sufridos, o se retiran y alejan del peligro para sacar duplicados de sí mismos. Tal es el juego que he dado en llamar "guerra nuclear". Se diferencia de casi todos los juegos de ordenador en que para nada intervienen personas en él. Los programas en liza, como es obvio, han sido escritos por personas; pero una vez trabados en batalla, sus creadores nada pueden hacer, sino observar inermes cómo el fruto de horas de diseño e implantación sobrevive, o muerde el polvo, en la pantalla del monitor. El resultado depende exclusivamente de cuál sea el programa que reciba un impacto en zona vulnerable.

La denominación "guerra nuclear" tiene su origen en una tecnología de construcción de memorias hoy abandonada. Durante los decenios de 1950 y 60, los sistemas de memoria de los ordenadores se componían de miles de núcleos ferromagnéticos anulares, enhebrados en una tupida red de finísimos hilos. Cada núcleo podía retener el valor de un dígito binario –un *bit*–, que es la unidad fundamental de información. En la actualidad, los elementos de memoria interna de los ordenadores se construyen sobre finas obleas de materiales semiconductores; no obstante, la parte activa del sistema de memoria, donde se encuentran almacenados los programas mientras se ejecutan, suele denominarse todavía memoria-núcleo, o simplemente, núcleo.

Los programas de combate de la "guerra nuclear" están escritos en un lenguaje especializado que hemos bautizado Redcode, pariente cercano de la familia de lenguajes de programación llamados ensambladores. En nuestros días, casi todos los programas de ordenador se redactan en lenguajes "de alto nivel", como los Pascal, Fortran o BASIC, que son lenguajes donde basta un enunciado para especificar toda una secuencia de instrucciones elementales de máquina. Además, al programador le resulta fácil leer y comprender los enunciados. Empero, para ejecutar un programa es preciso traducirlo primero a "lenguaje de máquina", donde cada instrucción queda representada por una larga secuencia de dígitos binarios. La redacción de programas directamente en esta forma es, en el mejor de los casos, tediosa.

Los lenguajes "ensambladores" ocupan posición intermedia entre los llamados lenguajes de alto nivel y el código de máquina. Redactadas en lenguaje ensamblador, cada una de las instrucciones de un programa corresponde, por lo general, a una única instrucción elemental y, por tanto, a una única cadena de dígitos binarios. Sin embargo, el programador, en lugar de escribir cadenas de unos y ceros, las representa mediante palabras breves, o abreviaturas *mnemónicas*, más fáciles de recordar que los números. La traducción de estas abreviaturas a lenguaje de máquina queda encomendada a un programa, llamado programa "ensamblador".

La proporción de programas redactados en lenguaje ensamblador es relativamente pequeña, porque los programas que generan resultan, al escribirlos, más largos y difíciles de comprender que sus homólogos de alto nivel. Hay, no obstante, ciertas tareas para las que los lenguajes ensambladores resultan ideales. Cuando un programa debe ocupar el mínimo espacio posible, o debe funcionar tan velozmente cuanto sea factible, se redacta en lenguaje ensamblador. Además, en lenguaje ensamblador pueden hacerse cosas del todo imposibles en lenguajes de alto nivel. Por ejemplo, puede hacerse que un programa modifique sus propias instrucciones, o se traslade a una nueva ubicación en la memoria.

La guerra nuclear vino sugerida por una anécdota que oí contar hace algunos años, acerca de un malévolo programador de un gran laboratorio de investigación de una multinacional, que llamaremos *X*. El programador redactó en lenguaje ensamblador un programa, Cizaña, que se duplicaba a sí mismo cada vez que era ejecutado. El programa era también capaz de difundirse de un ordenador a otro de la red de cómputo de la compañía; y no tenía más función que su propia perpetuación. Al poco, las copias de cizaña eran tan numerosas que otros programas y datos más útiles se asfixiaban y desalojaban por superpoblación. La creciente infección no pudo atajarse hasta que alguien pensó en combatir el fuego con el fuego. Se redactó un segundo programa autoduplicante, llamado Segador, cuyo propósito consistía en destruir copias de Cizaña, hasta que no pudiera encontrar más, y, seguidamente, destruirse a sí mismo. El Segador llevó a cabo su tarea, y las aguas volvieron a su cauce en el laboratorio de la gran compañía.

A pesar de las evidentes lagunas de la historia, yo la creí, seguramente porque así lo estaba deseando. Hizo falta algún tiempo para recomponer los verdaderos acontecimientos subyacentes a esta pieza del anecdotario informático. (Daré más adelante una reseña.) Basta por ahora hacer notar que mi deseo de creer descansaba plenamente en la intrigante idea de dos programas enzarzados en combate en los oscuros y silenciosos corredores del núcleo.

El año pasado decidí que, incluso si la historia resultaba no ser cierta, podría hacerse acontecer algo similar. Preparé una versión preliminar de la "guerra nuclear , y con la ayuda de David Jones, alumno de mi departamento de la Universidad de Ontario Occidental, la hice funcionar. Desde entonces hemos ido desarrollando y

perfeccionando el juego, hasta un nivel francamente interesante.

La guerra nuclear consta de cuatro componentes fundamentales: una matriz de memoria de 8000 posiciones, o ubicaciones; un lenguaje de tipo ensamblador, llamado Redcode; un programa ejecutor, designado MARS (acrónimo de *M*emory *A*rray *R*edcode *S*imulator), y la colección de programas de combate que han de batallar. Se instalan en la memoria dos programas de combate, en ubicaciones tomadas al azar, de modo que ninguno de los programas sepa dónde se encuentra el otro. El programa MARS va ejecutando las instrucciones de que constan los programas de combate según una variante sencilla de la técnica de compartición de tiempos, utilizada para asignar los recursos de un único ordenador entre gran número de usuarios. Los dos programas intervienen por turno: se ejecuta una primera instrucción del primer programa; después, una del segundo, y así sucesivamente.

Lo que haya de hacer un programa de combate durante los ciclos de ejecución que le sean asignados queda enteramente en manos del programador. El objetivo, como es obvio, es destruir el otro programa, reduciendo a ruinas las instrucciones que lo integran. También se admiten tácticas defensivas: un programa pudiera emprender la reparación de cualquier daño que haya recibido, o retirarse y hurtar el cuerpo al ser atacado. La batalla concluye en cuanto MARS llega en uno de los programas a una instrucción imposible de ejecutar. El programa al que corresponda tal instrucción defectuosa –presumiblemente, por heridas recibidas en combate– queda declarado perdedor.

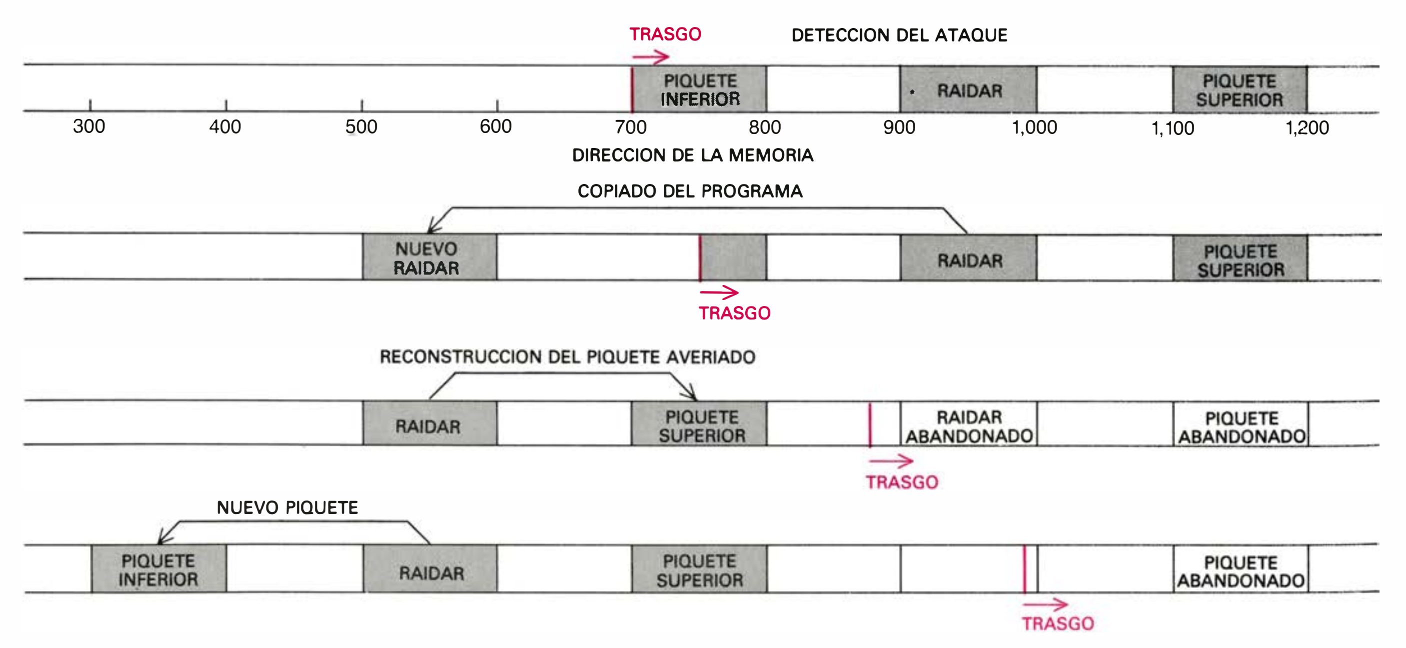
Mucho puede aprenderse acerca de un programa de combate sin más que analizar sus acciones, sea mentalmente o con lápiz y papel. No obstante, para someter el programa a la prueba de la experiencia es necesario acceder a un ordenador y disponer de una versión de MARS adecuada a él. Es factible hacer funcionar los programas en ordenadores personales; Jones y yo hemos preparado unas breves indicaciones para quienes tengan gusto por confeccionarse un campo de batalla propio donde jugar a guerra nuclear.

Antes de entrar en la descripción de Redcode y de algunos programas de combate sencillos, me agradaría añadir algo sobre la organización de la memoria. Aunque ya he hecho notar que consta de 8000 ubicaciones, este número nada tiene de mágico; matrices menores podrían servir perfectamente bien. La matriz de memoria se distingue de casi todas las utilizadas en ordenadores a causa de su configuración cíclica: es una sucesión de "direcciones" (en el sentido de dirección postal) numeradas de 0 a 7999, punto en el cual la sucesión se muerde la cola, de forma que la dirección 8000 es equivalente a la dirección 0. El programa MARS reduce siempre las direcciones mayores que 7999, tomando el resto de su división entre 8000. Por tanto, si un programa ordena atacar la posición 9378, MARS interpretará que tal dirección es la 1378.

Redcode es un lenguaje de tipo ensamblador, simplificado, construido expresamente. Dispone de instrucciones para trasladar el contenido de una posición de memoria a otra dirección, para alterar su contenido por procedimientos aritméticos y para transferir el control adelante y atrás dentro del programa. Mientras que el producto de un ensamblador auténtico está formado por una secuencia de códigos binarios, la forma mnemónica de las instrucciones de Redcode la traduce MARS en un gran número entero en expresión denaria, que seguidamente se almacena en la matriz de memoria; cada una de las ubicaciones de la memoria ha de poder albergar uno de tales enteros. Es también MARS quien interpreta como instrucciones esos enteros y quien se ocupa de realizar las operaciones en ellas indicadas.

Se da una lista de las instrucciones elementales de Redcode en la figura 2. El programador ha de suministrar a cada instrucción un argumento –un valor– cuando menos; casi todas las instrucciones requieren dos argumentos. Por ejemplo, en la instrucción JMP −7, la abreviatura JMP (de *jump*, saltar) va seguida del único argumento −7. La instrucción le dice a MARS que transfiera el control a la dirección de memoria situada siete lugares antes de la actual, esto es, siete lugares antes que la propia instrucción JMP −7. Si JMP −7 estuviera ocupando la dirección 3715, la ejecución del programa saltaría, retrocediendo hasta la dirección 3708.

Este procedimiento de cálculo de posiciones de la memoria se conoce por consignación (o "direccionamiento")



1. Por su refinada concepción, Raidar logra esquivar al Trasgo

INSTRUCCION	ABREV.	COD.	ARGUMENTOS		EXPLICACION
Trasladar (Move)	MOV	1	A	B	Traslada el contenido de la dirección *A* hasta la dirección *B*.
Sumar (Add)	ADD	2	A	B	Suma el contenido de la dirección *A* a la dirección *B*.
Restar (Substract)	SUB	3	A	B	Resta el contenido de la dirección *A* de la dirección *B*.
Saltar (Jump)	JMP	4	A		Transfiere el control a la dirección *A*.
Saltar si es cero (Jump if zero)	JMZ	5	A	B	Transfiere el control a la dirección *A* si el contenido de la dirección *B* es cero.
Saltar si es mayor (Jump if greater)	JMG	6	A	B	Transfiere el control a la dirección *A* si el contenido de la dirección *B* es mayor que cero.
Decrementar; saltar si es cero (Decrement; jump if zero)	DJZ	7	A	B	Resta 1 del contenido de la dirección *B* y transfiere el control a la dirección *A* si el contenido de la dirección *B* es entonces igual a cero.
Comparar (Compare)	CMP	8	A	B	Compara los contenidos de las direcciones *A* y *B*; si son distintos, pasa por alto la instrucción siguiente.
Declaración de dato (Data statement)	DAT	0		B	Enunciado no ejecutable. *B* es el valor del dato.

2. Lista de instrucciones de Redcode, lenguaje ensamblador para la guerra nuclear

relativa. El método de consignación relativa es el único utilizado con Redcode. No hay manera, pues, de que un programa conozca la posición absoluta propia que ocupa en la matriz de memoria.

La instrucción MOV 3 100 le indica a MARS que avance tres posiciones de memoria, copie lo que allí encuentre y lo deposite 100 posiciones más allá de la ocupada por la instrucción MOV, escribiendo sobre lo que allí pueda haber. Los argumentos de esta instrucción están dados en modo "directo", lo que significa que deberán interpretarse como direcciones sobre las que actuar. Se dispone de otros dos modos de direccionamiento. Al anteponer a un argumento el signo @, el argumento se convierte en "indirecto". En la instrucción MOV @ 3 100, el entero que ha de depositarse en la posición relativa 100 no es el que ocupa la posición relativa 3, sino el que se encuentre en la dirección especificada por el contenido de la dirección relativa 3. (La figura 4 da más detalles sobre el proceso de consignación indirecta.) Un signo # convierte al argumento en "inmediato", con lo que deja de tratarse como dirección para considerarse número entero. La instrucción MOV #3 100 tiene por efecto que el número 3 se desplace a la dirección relativa 100.

De las restantes instrucciones, la mayoría no requieren de ulterior explicación; pero la declaración de datos (DAT) sí precisa de algún comentario. Puede servir como zona de trabajo donde alojar información a la que un programa tal vez necesite referirse. Hablando estrictamente, toda localización de la memoria cuya primera posición decimal esté ocupada por un 0 puede considerarse una declaración DAT, y como tal, no ser ejecutable. Si a MARS se le pidiera ejecutar una tal "instrucción", sería incapaz de hacerlo, y declararía perdedor al correspondiente programa.

El entero denario en que se han codificado las instrucciones Redcode consta de varios campos, o áreas funcionales [*véase la figura 3*]. La primera cifra representa a la propia abreviatura mnemónica de la instrucción; dos cifras más identifican el modo de consignación (directa, indirecta, o inmediata). Se reservan, además, otras cuatro cifras para cada argumento. Los argumentos negativos se almacenan en forma de complementos: −1 se representaría como 7999, pues en la organización cíclica de la matriz de memoria, sumar 7999 tiene igual efecto que restar 1.

En la figura 5 se recogen las instrucciones que componen un sencillo programa de combate llamado Enano. El programa Enano es un programa tan temible como estúpido, que se abre paso a través de la memoria minando con un cero la quinta de cada cinco direcciones. El cero es el entero que denota enunciados de datos no ejecutables, por lo que un cero dejado caer en un programa enemigo puede detenerlo.

Supongamos que el programa Enano ocupe las posiciones absolutas números 1 a 4. La dirección 1 contiene inicialmente DAT −1, pero la ejecución comienza con la instrucción siguiente, ADD #5 −1. El efecto de esta instrucción es incrementar en el entero 5 el contenido de la dirección inmediatamente precedente, a saber, la declaración DAT −1, que de este modo quedará convertida en DAT 4. Seguidamente, el Enano ejecuta la instrucción de dirección absoluta 3, concretamente, MOV #0 @ −2. Aquí, el entero a trasladar es 0, especificado como valor inmediato. La dirección receptora se calcula indirectamente, como sigue. Primero, MARS descuenta dos posiciones de la dirección 3, con lo que se sitúa en la dirección 1. A continuación, examina el valor del dato allí contenido, que es 4, y lo interpreta como una dirección expresada relativamente a la posición actual; con otras palabras, cuenta hacia delante cuatro lugares, a partir de la dirección 1, y deposita un 0 en la dirección 5.

La última instrucción del Enano, JMP −2, tiene por misión crear un

ABREVIATURA	ARGUMENTO *A*	ARGUMENTO *B*	CODIGO DE OPERACION	DIGITO MODAL: ARGUMENTO *A*	DIGITO MODAL: ARGUMENTO *B*	ARGUMENTO *A*	ARGUMENTO *B*
DAT		−1	0	0	0	0000	7999
ADD	#5	−1	2	0	1	0005	7999
MOV	#0	@−2	1	0	2	0000	7998
JMP	−2		4	1	0	7998	0000

MODOS DE CONSIGNACION:			
	INMEDIATO	#	0
	DIRECTO		1
	INDIRECTO	@	2

3. Codificación de las instrucciones de Redcode mediante números denarios

bucle infinito. Encamina nuevamente la ejecución a la dirección absoluta 2, la cual manda incrementar en 5 unidades la declaración DAT, con lo que su nuevo valor será DAT 9. En el nuevo ciclo de ejecución quedará depositado, pues, un 0 en la dirección absoluta 10. Posteriormente, irán cayendo, a modo de bombas, otros tantos ceros en las direcciones 15, 20, 25, etcétera. Aunque el programa propiamente dicho permanece inmóvil, su artillería puede batir de extremo a extremo la formación de memoria. Finalmente, el Enano termina por abrirse paso hasta las direcciones 7990, 7995, y al cabo, 8000. En cuanto a MARS concierne, 8000 es idéntico a 0, y como vemos, el Enano se libra por pelos del suicidio. Su proyectil siguiente vuelve a aterrizar en la dirección 5.

Los posibles entusiasmos iniciales quedan atemperados tan pronto nos damos cuenta de que ningún programa estacionario que conste de más de cuatro instrucciones podrá librarse de ser alcanzado por el Enano. El programa adversario tiene tan sólo tres opciones: irse moviendo de acá para allá, esquivando el bombardeo, encajar los impactos, reparar los daños sufridos o cazar él antes al Enano. Para que esta última estrategia sea de efecto no estará de más un poco de suerte: al no tener idea de dónde se encuentra residiendo el Enano, habrá de transcurrir por término medio 1600 ciclos antes de que se produzca un impacto. Si el segundo programa es también un Enano, cada programa vence un 30 por ciento de las veces; en el 40 por ciento de los combates, ninguno llega a golpear al otro con un impacto mortal.

Antes de tomar en consideración las otras dos estrategias, quisiera presentar un curioso programa de combate, que consta de una sola línea, que llamaremos Trasgo. Hélo aquí:

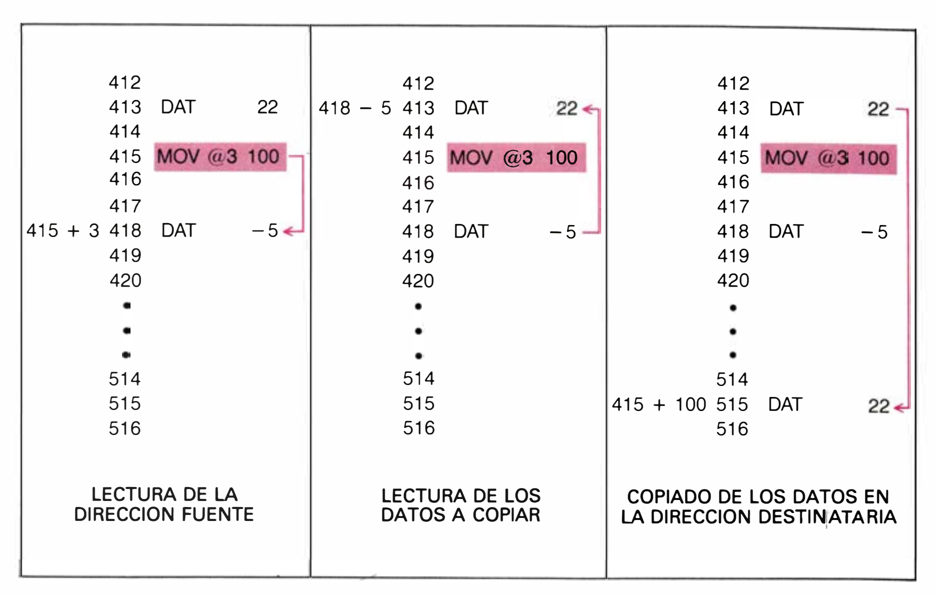
MOV 0 1

Tenemos con el Trasgo el ejemplo más sencillo de programa Redcode capaz de ir cambiando de posición por sí mismo. El programa copia el contenido de la dirección relativa 0 (a saber, MOV 0 1) en la dirección relativa 1, que es la siguiente a la que actualmente ocupa. Al ejecutar el programa, éste va desplazándose por la matriz, a razón de una dirección por ciclo, dejando tras de sí una estela de instrucciones MOV 0 1.

¿Qué sucederá si enfrentamos al Trasgo con el Enano? La barrera de ceros proyectada por el Enano avanza más rápidamente que el Trasgo a través de la matriz de memoria, pero de ello no se deduce que forzosamente haya el Enano de llevar ventaja. La cuestión es: ¿logrará el Enano hacer impacto en el Trasgo, incluso aunque haya llegado a elevarse la barrera?

Si el Trasgo es el primero en alcanzar al Enano, con toda probabilidad surcará y atravesará el código del Enano. Cuando la instrucción JMP −2 del Enano transfiera el control dos pasos hacia atrás, la instrucción que allí se encontrará será la MOV 0 1 del Trasgo. En consecuencia, el Enano quedará subvertido, convirtiéndose en un segundo Trasgo que perseguirá indefinidamente al otro, dando una y otra vuelta a la matriz de memoria. Con las reglas de la guerra nuclear, la batalla es un empate. (Obsérvese que éste es el resultado previsible "con toda probabilidad". Se invita a los lectores a analizar otras posibilidades, con lo que tal vez descubran el chocante resultado de una de ellas.)

Tanto el Enano como el Trasgo son representantes de una categoría de programas que podemos caracterizar como diminutos, muy agresivos, pero no inteligentes. En el nivel siguiente tenemos programas más extensos, y algo menos agresivos, pero lo bastante astutos como para habérselas con los de la categoría inferior. Los más perspicaces gozan de cierta capacidad de esquivar ataques, refugiándose en copias de sí mismos que los libren del peligro. Cada uno de tales programas habrá de contener un segmento de código similar al



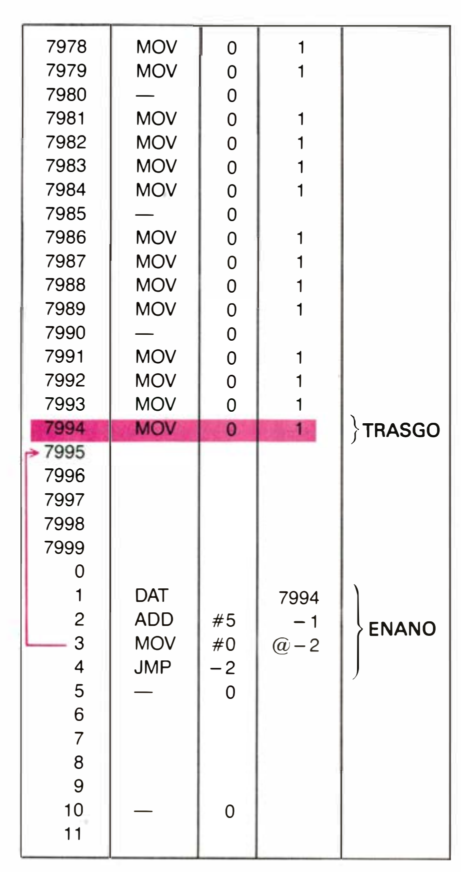
4. Mecanismo en tres pasos para la consignación relativa indirecta

DIRECCION	CICLO 1.°			CICLO 2.°			CICLO 9.°		
0									
1	DAT		−1	DAT		4	DAT		14
2	ADD	#5	−1	ADD	#5	−1	ADD	#5	−1
3	MOV	#0	@−2	MOV	#0	@−2	MOV	#0	@−2
4	JMP	−2		JMP	−2		JMP	−2	
5				—	0		—	0	
6									
7									
8									
9							—	0	
10									
11									
12									
13									
14							—	0	
15									
16									
17									

5. Enano es un programa de combate que lanza "barreras de ceros"

llamado Gemini, que podemos ver en la figura 7. El programa Gemini no está concebido para ser un programa de combate completo. Su única función es producir una copia de sí mismo 100 unidades más allá de su posición actual y transferir, después, el control a la copia nueva.

El programa Gemini consta de tres partes principales. Dos declaraciones de datos situadas en su principio hacen papel de indicadores: señalan la siguiente instrucción que habrá de copiarse y su destino. Un bucle, que ocupa la sección central del programa, se encarga de efectuar la copia, trasladando por turno cada instrucción hasta una dirección situada 100 lugares más allá de su actual ubicación. En cada tránsito por el bucle ambos indicadores se incrementan en 1, señalando así una nueva dirección fuente y una nueva dirección destino. Una instrucción de comparación (*CMP*), inserta en el bucle, va inspeccionando el valor de la primera declaración de datos; cuando tal valor se ha incrementado nueve veces, el programa entero ha sido copiado, y se toma una salida del bucle. Queda por efectuar un último reajuste. La dirección destinataria es la segunda



6. Trasgo se enfrenta a Enano. ¿Quién vencerá?

declaración de datos del programa; tiene un valor total inicial de 99. Pero en el momento de copiarse ha sido incrementada ya una vez, con lo que la nueva versión del programa dice DAT 100. La instrucción MOV #99 93 corrige el error de transcripción y seguidamente la ejecución se transfiere a la nueva copia.

Modificando Gemini puede crearse toda una nueva clase de programas de combate. Uno de ellos, el llamado Visnú, como la temible deidad india, se copia a sí mismo 10 lugares más adelante, en lugar de los 100 de Gemini. Al igual que en Trasgo, su propósito es arrollar a sus oponentes. Mas, a diferencia de aquél, gana con mucha más frecuencia y empata mucho menos, porque es mucho menos probable que un programa sobre-escrito a Visnú consiga ejecutar del código de éste. Otro programa que se vale del mecanismo de Gemini es el Patazas, que deja como intervalo entre copias un número primo grande. Patazas es difícil de atrapar, y produce sobre los códigos del programa enemigo efectos tan devastadores como Visnú.

Ni Visnú ni Patazas son muy inteligentes. Hasta ahora sólo hemos escrito dos programas que puedan aspirar al segundo grado de perfeccionamiento y cualificación. Ambos son demasiado largos para reproducirlos aquí. Uno de ellos, que llamamos Raidar, mantiene dos "piquetes de guardia" que protegen los flancos del programa propiamente dicho [*véase la figura 1*]. Cada piquete consta de 100 direcciones consecutivas, ocupadas por "unos", separadas del programa principal por un foso o amortiguador de 100 direcciones vacías. Raidar reparte su tiempo entre ataques sistemáticos a zonas distantes de la formación de memoria y la revisión de sus piquetes de protección. Caso de descubrir alterado alguno de los puestos de vigilancia, Raidar interpreta la novedad como indicación de haber sufrido el ataque de un Enano, de un Trasgo, o de algún otro programa sin inteligencia. Raidar saca entonces una copia de sí mismo hacia el otro lado del piquete dañado, lo reconstruye, crea un piquete nuevo que proteja su flanco vulnerable y reanuda su funcionamiento normal.

Además de la capacidad de sacar copias de sí mismos, podemos dotar a los programas de combate de capacidad de autorreparación. Jones ha preparado un programa autorregenerativo capaz de sobrevvir a ciertos ataques, si bien no a todos. Llamado Inspector, conserva dos copias de sí mismo, si bien ordinariamente sólo se ejecuta una de ellas. La copia que se encuentre en funcionamiento examina periódicamente la otra, para comprobar si alguna de las instrucciones ha quedado alterada por cualquier ataque. La detección de cambios se consigue comparando ambas copias, suponiendo siempre que la copia en ejecución es la correcta. Cuando se detectan cualesquiera instrucciones anormales se las reemplaza, y se transfiere el control a la segunda copia, que comienza entonces a examinar el estado de la primera.

Siento curiosidad por saber si los lectores consiguen idear otros tipos de programas autorreparadores. Por ejemplo, podríamos pensar en mantener dos o más ejemplares de un mismo programa, aun cuando llegue a ejecutarse tan sólo una copia. El programa podría dar cabida a una sección encargada de las reparaciones, que para regenerar las instrucciones deterioradas recurriera a una copia supernumeraria. La sección de reparación podría incluso repararse a sí misma, si bien puede que siga siendo vulnerable al ser atacada en ciertas posiciones. Una medición de vulnerabilidad considera que ha sido alcanzada únicamente una instrucción. Por término medio, ¿cuántas de tales instrucciones, de ser alcanzadas, llegan finalmente a producir la muerte del programa? Con este criterio de medición, ¿cuál es programa autorreparador menos vulnerable que podemos escribir?

Tan sólo si se logra desarrollar programas razonablemente robustos llegará la "guerra nuclear" a resultar juego interesante, en el que se cargue más el acento en el aspecto ofensivo que en el defensivo. Los programas de combate tendrán entonces que buscar e identificar los códigos del enemigo y saber desencadenar ataques concentrados donde los descubran.

Tal vez haya yo dado la impresión de que el sistema formado por el lenguaje Redcode y el programa MARS están fijados de una vez por todas. No es así. En ratos libres hemos estado experimentando nuevas ideas; estamos, desde luego, abiertos a sugerencias. De hecho, tanto hemos experimentado con nuevos programas y nuevas características que hay todavía batallas por librar en nuestro sistema.

Una idea con la que hemos estado jugueteando consiste en incluir una instrucción adicional que facilitara

	DAT		0	/indicador de dirección fuente
	DAT		99	/indicador de dirección destino
	MOV	@ −2	@ −1	/copiado de la fuente en el destino
	CMP	−3	#9	/si las 10 líneas han sido copiadas...
BUCLE	JMP	4		/... salir del bucle;
	ADD	#1	−5	/de lo contrario, incrementar la dirección fuente...
	ADD	#1	−5	/... y la dirección destino...
	JMP	−5		/... y regresar al bucle
	MOV	#99	93	/reinicializar la dirección fuente
	JMP	93		/saltar a la copia nueva

7. Gemini, un programa que saca copias de sí mismo en nuevas ubicaciones de la matriz de memoria

un poco la autoprotección o la autorreparación. La instrucción PCT *A* serviría para preservar de alteraciones la instrucción emplazada en la dirección *A* hasta la siguiente vez en que se ejecutara. ¿En qué medida podríamos disminuir la vulnerabilidad de un programa sacando partido de una instrucción de ese tipo?

En las directrices de que hablábamos al principio no sólo describimos las reglas de la guerra nuclear, sino que explicamos también cómo preparar una matriz de memoria y redactar un programa MARS en diversos lenguajes de alto nivel. Sugerimos, también, posibles formas de presentar los resultados de las batallas. Por el momento, las siguientes reglas definen el juego con suficiente precisión para que los aficionados que dispongan solamente de lápiz y papel puedan comenzar el diseño de programas de combate:

1. Los dos programas de combate se cargan en posiciones tomadas al azar en la matriz de memoria, inicialmente separadas por mil direcciones, cuando menos.
2. MARS va ejecutando alternativamente una instrucción de cada programa, hasta tropezar con una instrucción que no pueda ejecutarse. El programa que contiene la instrucción errónea pierde.
3. Los programas pueden ser atacados con cualesquiera armas disponibles. Las "bombas" pueden ser ceros, otro entero o cualesquiera instrucciones válidas de Redcode.
4. A cada ronda se le impone una duración límite, determinada por la velocidad del ordenador. Si, alcanzado el límite de tiempo, ambos programas siguen en funcionamiento, el combate se declara nulo.

La anécdota de la Cizaña y el Segador parece estar basada en un refrito de dos programas verdaderos. Uno de ellos era un juego de ordenador llamado Darwin, inventado por M. Douglas McIlroy, de los American Telephone & Telegraph Bell Laboratories. El otro se llamaba Worm (Gusano) y fue redactado por John F. Shoch, del Centro de Investigación de Xerox en Palo Alto. Ambos programas tienen ya algunos años, lo que ha permitido holgadamente que se propalasen todo tipo de rumores. (Darwin se describe en *Software: Practice and Experience*, volumen 2, págs. 93-96, 1972. Se da también una vaga descripción de lo que parece ser el mismo juego en la edición de 1978 de *Computer Lib*.)

En el juego Darwin, cada jugador presenta cierto número de programas redactados en lenguaje ensamblador, llamados "organismos", que habitan en la memoria central conjuntamente con los organismos presentados por los demás contendientes. Los seres creados por cada jugador (pertenecientes por tanto a una misma "especie") tratan de aniquilar a los de otras especies que ocupan su espacio. Gana la partida el jugador cuyos organismos sean más abundantes al terminar el tiempo de combate. McIlroy inventó un organismo invulnerable, que "ganó tan sólo unos pocos juegos". Al parecer, aunque inmortal, no era muy agresivo.

El Gusano era un programa experimental ideado para sacar el máximo partido de una red de miniordenadores interconectados de Xerox. Un programa supervisor se encargaba de cargar el Gusano en máquinas inactivas; su finalidad era asumir el control de la máquina y, en combinación con Gusanos residentes en otras máquinas inactivas, hacer funcionar grandes programas de aplicación en el sistema multiprocesador resultante. Gusano se diseñó de modo que quienquiera desease utilizar una de las máquinas por él ocupadas pudiera reclamarla fácilmente sin interferir el trabajo principal.

Podemos detectar elementos tanto de Darwin como del Gusano en la historieta de la Cizaña y el Segador. En la guerra nuclear, el Segador ha cobrado realidad.

Taller y laboratorio

Al calentar un alambre se revelan los cambios que atraviesa la estructura cristalina del acero

Jearl Walker

Son numerosas las propiedades aprovechables del hierro y del acero que dependen de la estructura cristalina del metal. Sutiles variaciones de ésta pueden ser causa de alteraciones radicales en sus propiedades como consecuencia de un calentamiento o de un enfriamiento. Charles F. Walton, metalúrgico e ingeniero mecánico de Cleveland, ha ideado un ensayo para evidenciar cuán profundamente cambia con el calor un acero ordinario, que es una aleación binaria, o sea, formada por dos componentes. Aunque este experimento es de realización muy sencilla, presenta varios motivos de asombro, de los cuales no todos están suficientemente explicados.

Walton calienta y enfría un trozo de 150 centímetros de cuerda de piano (alambre de acero del n.º 29). La experiencia la lleva a cabo sujetando horizontalmente el alambre entre dos bornes montados en sendos bloques de madera y conectándolo a un variac; éste le permite someter al alambre a corrientes eléctricas de intensidad controlada. Rápidamente sube el variac hasta unos 55 volt, con lo que alimenta el alambre con unos 14 ampère. (Esta corriente excede con mucho el límite del variac, pero se mantiene durante un tiempo demasiado corto para que perjudique al aparato.)

La corriente eléctrica calienta el alambre hasta el punto de que éste se dilata, se afloja y no tarda en ponerse al rojo. Cuando corta la corriente, el color rojo palidece y el alambre se contrae. Y aquí se plantea el primer enigma: durante un instante el alambre enrojece de nuevo y torna a relajarse. Luego, prosigue enfriándose. Por alguna causa, el alambre, durante su enfriamiento, libera energía en un brusco centelleo rojo. ¿De dónde procede tal energía? ¿Por qué no se libera continuamente a lo largo del enfriamiento?

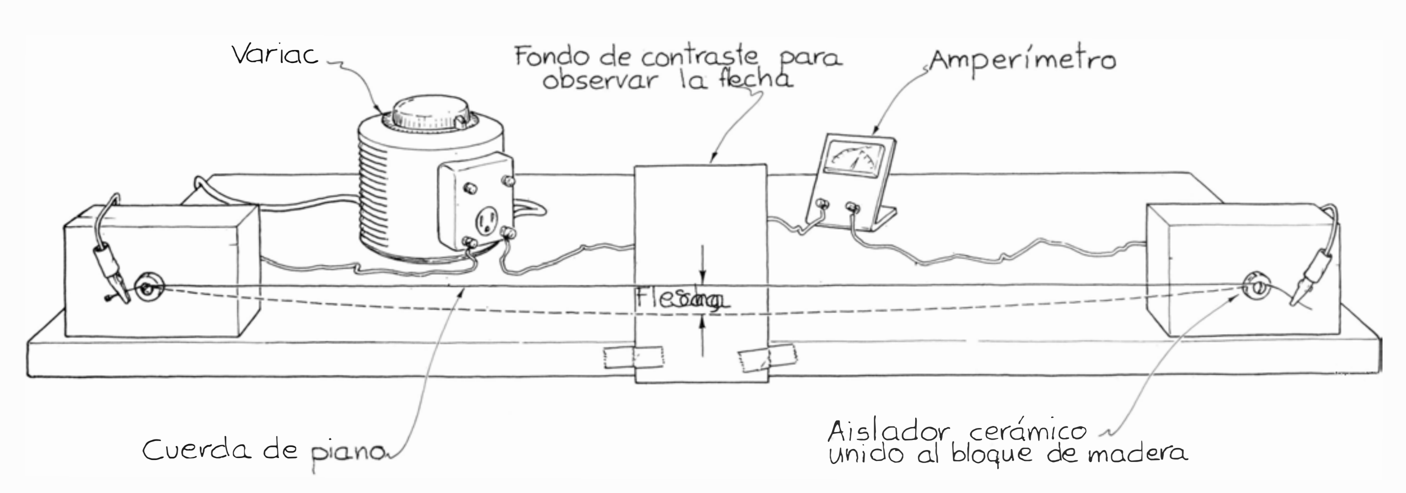
El segundo enigma está relacionado con la velocidad de enfriamiento del alambre. En el curso de sus experiencias, Walton recalentó el alambre, interrumpió la corriente y envolvió, durante unos segundos, una porción del alambre al rojo vivo con una esponja húmeda. El agua enfrió enseguida esa porción, pero el resto del alambre tardó un rato más en enfriarse. La parte rápidamente enfriada resultó tan frágil que se quebró sin resistencia entre sus dedos. Además, la punta del trozo roto tenía una dureza suficiente para arañar el vidrio. La parte lentamente enfriada no se quebraba con facilidad, ni la punta de un trozo separado de la misma arañaba el vidrio. ¿Qué le pasa a la porción prestamente enfriada para que se alteren su ductilidad y dureza?

El tercero de los enigmas se le presentó a Walton respecto a las propiedades magnéticas del alambre. A la temperatura ambiente, un pequeño imán que se sitúe próximo a él lo atraerá. Pero, cuando está al rojo, no se observa en el alambre reacción apreciable a la presencia del imán. ¿Por qué puede depender de la temperatura el magnetismo de un material? De hecho, ¿qué hay en el alambre frío que responda de la atracción magnética?

A quien se proponga repetir sus experimentos Walton sugiere que se haga con alambre de la misma medida (unos 2 milímetros de diámetro). Si el alambre es más delgado, al calentarse se oxidará tanto que servirá sólo para dos o tres ensayos. Si es más grueso necesitará una corriente demasiado intensa para calentarse. Hay que extremar el cuidado para no tocar el alambre mientras lo atraviese la corriente. ¡*Peligro de muerte*! Antes de aplicar el imán o la esponja húmeda se desconectará o desenchufará el variac.

Estos enigmas de Walton podemos desvelarlos ocupándonos primero del hierro, constituyente principal del acero. El hierro es alotrópico, lo que significa que puede presentar diferentes formas cristalinas en estado sólido. Un cristal suele describirse en función de la ordenación de la mínima unidad posible de sus átomos. El resto del cristal no es sino una repetición de esta celda unidad. A la temperatura ambiente, la celda unidad del hierro consta de un átomo ubicado en el centro de un cubo formado por otros ocho átomos en cada vértice. Esta configuración recibe el nombre de estructura cúbica centrada en el cuerpo. Y el hierro con los átomos así dispuestos se conoce como hierro alfa o ferrita.

Habitualmente, el hierro se compone



1. Montaje de Charles F. Walton para calentar cuerdas de piano

de un gran número de regiones cristalinas independientes llamadas granos. Dentro de éstos, las celdas unidad están orientadas uniformemente, pero los granos están orientados al azar. Los granos se forman al enfriarse el hierro caliente, y los cristales comienzan a formarse en los puntos de nucleación, o de germinación, continuando su crecimiento hasta que acaban poniéndose en contacto entre ellos, para formar una matriz de granos.

Cuando el hierro se calienta hasta los 910 grados Celsius, la estructura de sus cristales cambia de la cúbica centrada en el cuerpo a la cúbica centrada en las caras, característica del hierro gamma, o austenita. Esta celda cúbica posee átomos en los vértices y en los centros de las caras. Tenemos asimismo granos, compuestos cada uno por un cristal bajo una orientación dada.

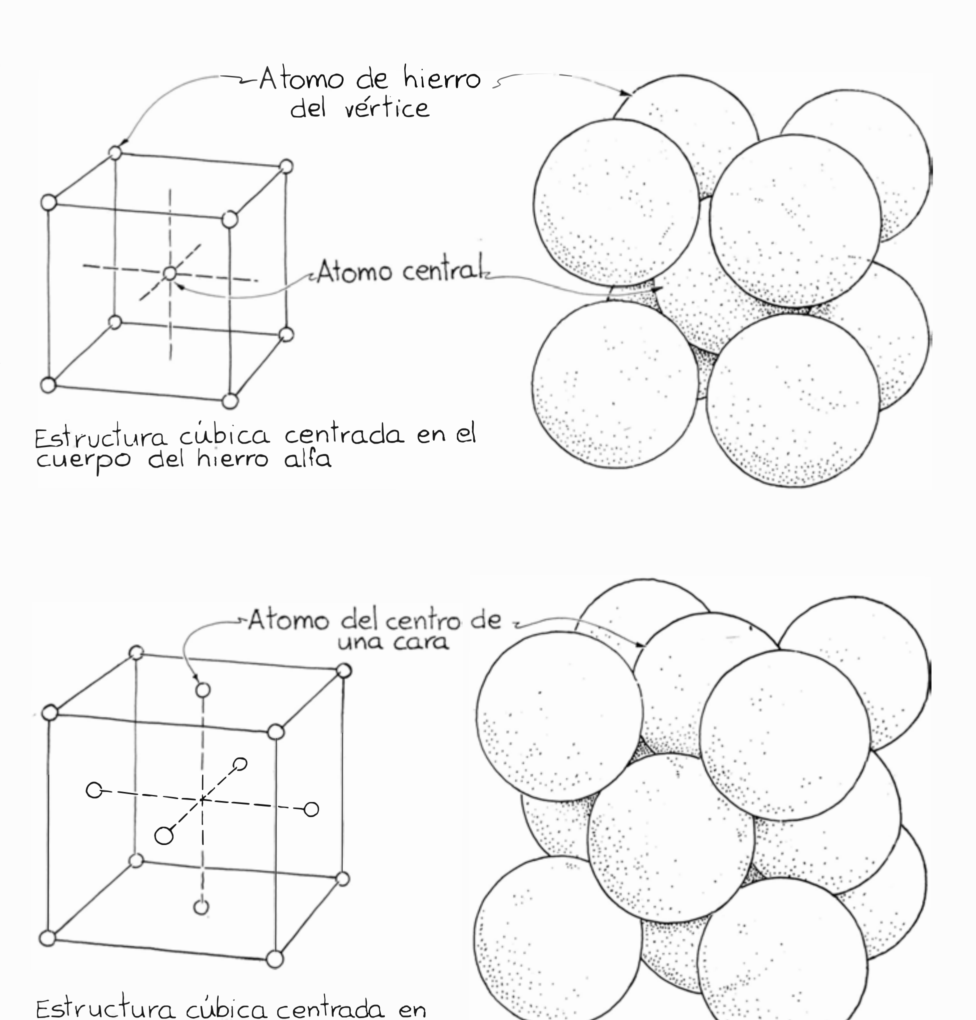
El paso de hierro alfa a hierro gamma requiere energía para reordenar los átomos en la nueva estructura. Otra forma, llamada hierro delta, aparece a una temperatura mucho más alta que las de las experiencias de Walton. Si el hierro se calienta todavía más, se funde.

Cuando se calienta el hierro alfa, toda aportación de energía se invierte al principio en elevar la temperatura. Pero, alcanzado el punto de transición, la temperatura debe permanecer constante hasta que se haya aportado energía suficiente para que todos los cristales hayan pasado a hierro gamma. Mientras tanto, el calentamiento no producirá aumento de la temperatura.

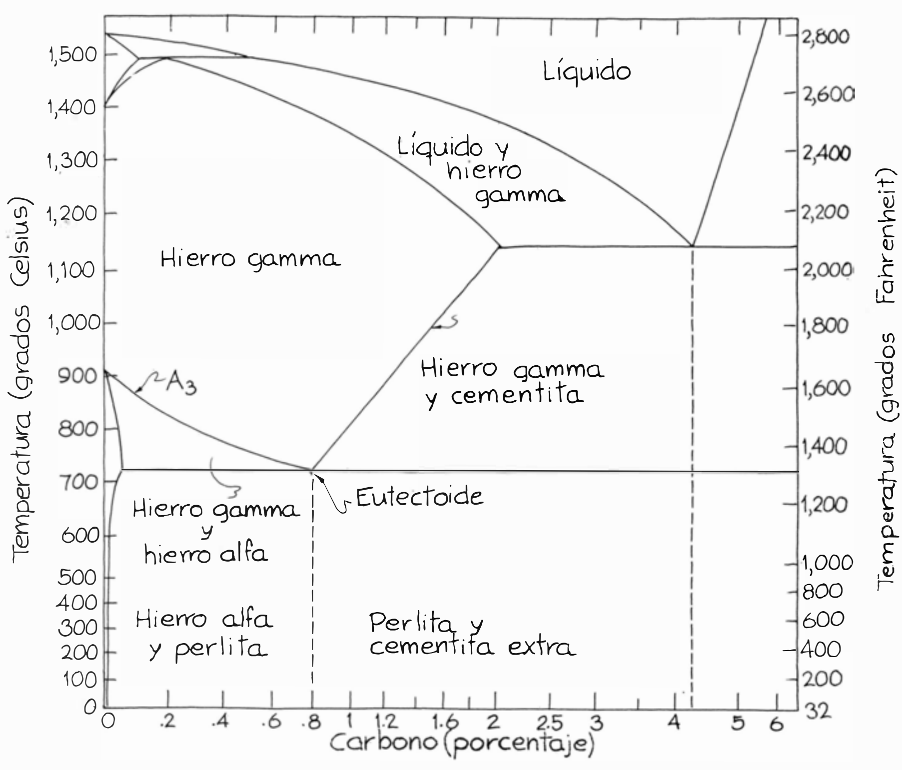
Ocurre también lo contrario. Cuando se enfría el hierro gamma, su temperatura desciende hasta el punto de transición. Entonces habrá que eliminar calor hasta que los cristales hayan regresado al estado alfa. Mientras tanto, la temperatura no podrá comenzar a bajar de nuevo.

El punto de transición entre el hierro alfa y el gamma es similar al punto de solidificación y de fusión del agua. Así, cuando se calienta hielo, su temperatura aumenta hasta llegar al punto de fusión y no puede aumentar más hasta que se haya derretido todo. Cuando se enfría el agua, ésta debe permanecer en el punto de solidificación hasta que haya congelado toda. Mientras tanto, su temperatura no podrá bajar.

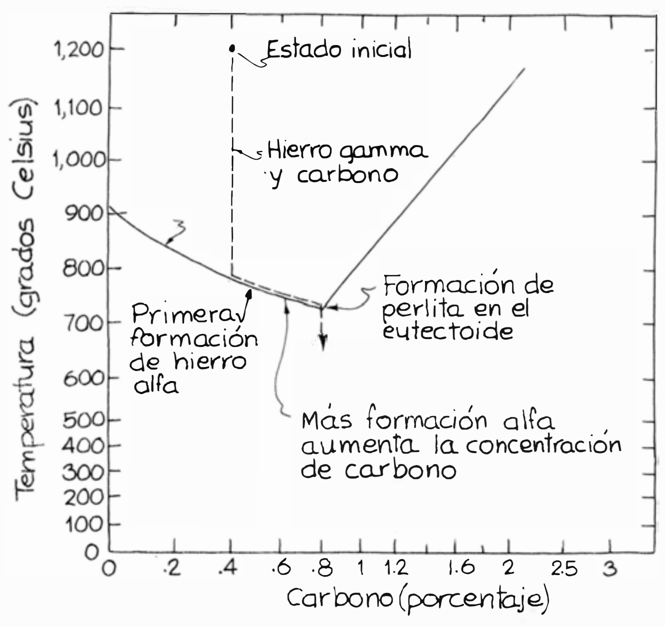
La cuerda de piano de que se vale Walton se compone esencialmente de hierro alfa. Cuando la corriente la atraviesa, los choques de los electrones que la integran con la estructura cristalina generan calor. Por ello, el alambre acaba convirtiéndose en hierro gamma.



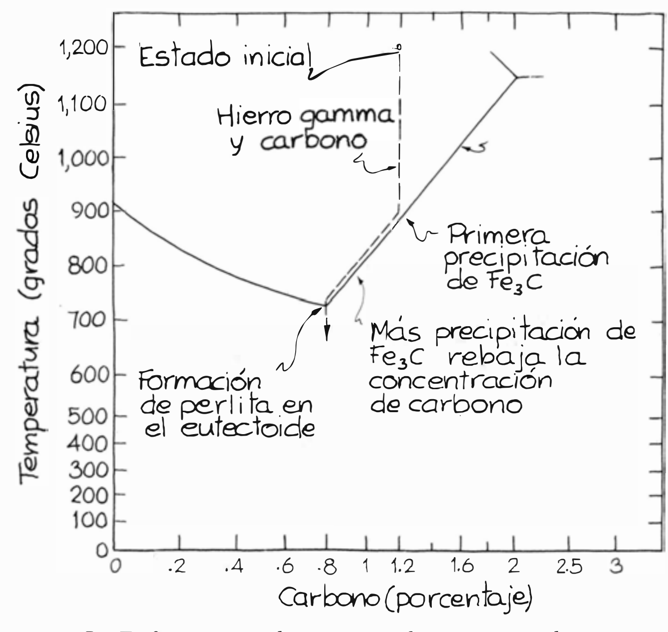
2. Estructura de la celda unidad en dos tipos de cristal de hierro



3. Diagrama de fases de una aleación de hierro y carbono



4. *Enfriamiento de un acero hipoeutectoide*



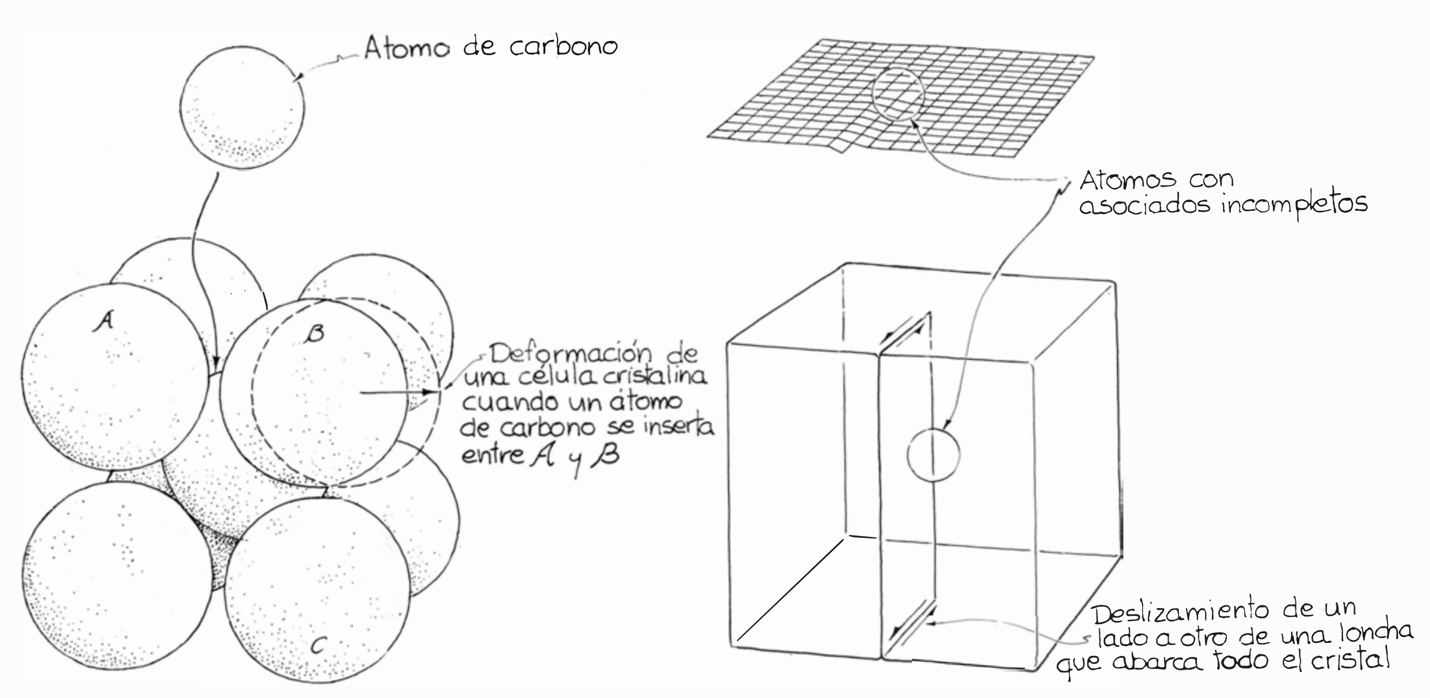
5. *Enfriamiento de un acero hipereutectoide*

Y el calor adicional lo calienta tanto que no tarda en emitir en la zona roja del espectro visible.

Cuando se corta la corriente, el alambre se enfría y se atenúa la emisión visible. A la temperatura de transición, el reordenamiento de los cristales cúbicos centrados en las caras al pasar a centrados en el cuerpo libera energía, la cual recalienta momentáneamente el alambre hasta el punto de que éste enrojece y se afloja otra vez. Este fulgor es breve, pues esa energía se pierde rápidamente por radiación y convección en el aire. O sea, el corto centelleo rojo observado en el experimento de Walton corresponde a la energía que se libera en la transición de hierro gamma a hierro alfa. Constituye, además, la prueba de que la ordenación de los átomos de una de las formas cristalinas requiere más energía que la ordenación de los átomos de la otra.

Para correr el velo de otro de los misterios de los ensayos de Walton debemos tener en cuenta que el acero contiene carbono. Por supuesto, existen numerosas aleaciones de acero, pero aquí me refiero únicamente a la binaria de hierro y carbono. El análisis nos lo facilitará un diagrama de fases del tipo que se ilustra en la figura 3. En éste, las ordenadas representan las temperaturas de la aleación y, las abscisas, el porcentaje de carbono en el hierro.

Con el acero en estado líquido, el carbono se disuelve sin dificultad en el hierro. Pero incluso con el metal en estado sólido el carbono puede entremezclarse con los cristales de hierro para formar lo que se llama una solución sólida. La solubilidad del carbono en



6. *Un átomo de carbono deforma la celda unidad*

7. *Dislocación en borde de un cristal*

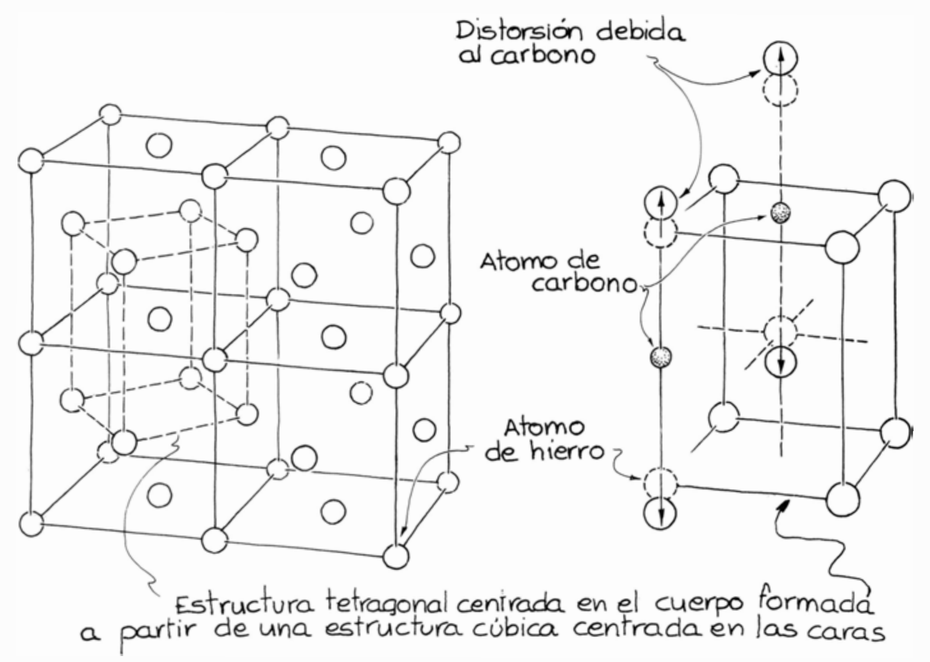
dicha solución desempeña un papel muy importante en el experimento de Walton.

En el segundo diagrama de fases [*véase la figura 4*] se señala un punto en el que el alambre se halla en estado gamma a una temperatura de 1200 grados Celsius para un contenido de carbono de 0,4 por ciento, o sea, con cuatro partes de carbono por cada 1000 partes de solución sólida. Este carbono está disuelto en el hierro gamma en el sentido de hallarse disperso por toda la estructura cristalina, de tal modo que sus átomos se han encajado forzadamente en las aristas de las celdas unidad. Pero sólo unas pocas de las aristas alojan carbono, ya que el contenido de éste es reducido.

En el diagrama, el enfriamiento de este acero se representa mediante una recta vertical que se extiende desde el punto inicial hasta una línea rotulada A_3. Esta jalona la transición de hierro gamma a hierro alfa. Hasta que se alcanza este punto de transición, toda extracción de energía del hierro reducirá la temperatura, pero no variará ni la estructura cristalina ni la solubilidad del carbono en tal estructura. Una vez en A_3 la eliminación de energía siguiente obligará a que una parte del hierro gamma (principalmente el de los límites de grano) pase a hierro alfa.

El carbono es casi insoluble en el hierro alfa (el límite de solubilidad es sólo del 0,025 por ciento), por lo que la solución de hierro gamma y carbono se hará más concentrada. En el diagrama de fases ello se traducirá en que el trayecto de enfriamiento a partir de A_3 sigue hacia abajo y hacia la derecha, pues a más formación de hierro alfa corresponde mayor concentración de carbono en el hierro gamma restante. Mientras que en el hierro puro la transición entre hierro gamma y hierro alfa tiene lugar a una única temperatura, la presencia del carbono despliega el proceso en una gama de temperaturas.

Finalmente, a la temperatura de 723 grados Celsius, se llega al extremo de A_3. Este estado, llamado eutectoide, marca la máxima concentración posible (0,8 por ciento) de carbono en hierro gamma. Toda eliminación ulterior de energía obliga a la solución que queda a precipitarse en capas alternas de hierro alfa y conglomerados de carburo de hierro (Fe_3C), comúnmente llamado cementita. Esta combinación de hierro alfa y cementita se llama perlita. Dado que un mayor enfriamiento no altera la mezcla, el acero es perlítico a la temperatura ambiente.



8. Estructura cristalina de la martensita

Al calentar el acero ese proceso se invierte. El metal será perlítico hasta llegar al eutectoide. Con más calor, la cementita comenzará a transformarse en una solución sólida de hierro gamma y carbono. Con más calor aún, los restos de hierro alfa comenzarán a transformarse en hierro gamma y disminuirá la concentración de carbono en éste. En el diagrama de fases el trayecto será ascendente y hacia la izquierda, a lo largo de A_3. Cuando ya no quede hierro alfa, el trayecto abandonará la línea A_3 y se dirigirá hacia arriba, paralelamente al eje de temperaturas. En este período del calentamiento, permanecerán constantes la estructura cristalina y la concentración de carbono.

Cuando un acero en estado gamma tiene una concentración de carbono inferior al 0,8 por ciento, propia del eutectoide, del mismo se dice que es hipoeutectoide. Los aceros hipereutectoides, cuya concentración de carbono es superior al 0,8 por ciento, se enfrían casi del mismo modo que les he descrito, salvo que inicialmente precipitan carburo de hierro y no hierro alfa.

Tomemos un acero en estado gamma con una concentración de carbono del 1,2 por ciento. Cuando se enfríe el metal, dicha concentración permanecerá constante hasta llegar a un punto de transición. En nuestro tercer diagrama de fases [*véase figura 5*] la línea que nos importa ahora está rotulada A_{cm}.

Si la muestra se enfría más, no todo el carbono podrá permanecer en solución en el hierro gamma. El enfriamiento obligará a que parte del carbono (principalmente en los límites de grano del hierro gamma) precipite en carburo de hierro. Así disminuirá la concentración de carbono. En el diagrama de fases el trayecto de enfriamiento sigue por A_{cm} hacia el eutectoide, donde más enfriamiento dará como resultado la formación de perlita. Cuando este acero llegue a la temperatura ambiente, tendrá mucho carburo de hierro mezclado con la perlita.

En estado gamma, el acero de cuerda de piano contiene alrededor del 0,8 por ciento de carbono. Este alambre, tras haber sido calentado a la temperatura del rojo, se enfría con suficiente rapidez para subenfriar la temperatura del eutectoide hasta los 550 o 600 grados Celsius. Entonces, se transforma bruscamente de carbono y hierro gamma en perlita, liberando energía. Por eso despide un destello rojo y se relaja momentáneamente. Parte de esa relajación resulta de la dilatación térmica cuando el alambre vuelve a calentarse a causa de la transformación de los cristales. El resto de la relajación débese a la dilatación inherente al reordenamiento de los átomos al formar perlita.

La solubilidad del carbono en el hierro depende primordialmente del espacio disponible en las celdas unidad. En la estructuración cúbica centrada en el cuerpo, propia del hierro alfa, un átomo de carbono puede colocarse en una arista del cubo o en el centro de una cara. Como el espacio disponible en ambas posiciones es inferior al tamaño del átomo de carbono, este debe desplazar de su posición correcta a un

átomo de hierro situado en un vértice. En la figura 6 se muestran dos posiciones posibles para un átomo de carbono intercalado en una arista. Si éste se abre paso entre los átomos de hierro *A* y *B* moviendo a *B*, *B* se desplazará de su posición hacia la derecha. Si, en otro caso, el átomo de carbono se abre paso entre *B* y *C*, *B* puede verse forzado hacia arriba. Cualquiera que sea el caso, el desplazamiento de *B* deformará notablemente la estructura cristalina. El carbono es esencialmente insoluble en hierro alfa porque dispone de poco espacio y su presencia deforma los cristales.

En el hierro gamma, el carbono está constreñido a las aristas de la celda cristalina. Aunque aquí los espacios para que anide el carbono son menos que en el hierro alfa, los existentes en las aristas son levemente mayores. La presencia del carbono crea, por tanto, menos distorsión que en el hierro alfa, lo que permite que se disuelva más.

El objeto fundamental del carbono en el acero es reforzar las propiedades mecánicas del metal. Los granos de hierro contienen casi siempre dislocaciones que interrumpen la ordenación regular de los cristales ideales y, por ello, debilitan los granos. El carbono consolida los granos anclando en su sitio a las dislocaciones.

Un tipo de irregularidad muy corriente es la dislocación en borde. Consideremos un cubo uniforme de hierro alfa con cada una de sus celdas interiores enlazada a celdas contiguas. Practiquemos una cisión imaginaria hasta la mitad del bloque y forcemos ambos labios de la cisión en sentidos contrarios una distancia igual al ancho de una celda. El cubo contendrá así una fila de átomos desalineados respecto a las celdas circundantes.

Estas dislocaciones en borde debilitan el grano. Efectivamente, si en estas condiciones se aplica una tensión cortante de tal modo que haga deslizar aún más los bloques cristalinos, la línea de dislocación vertical se abrirá paso fácilmente a través del cristal. Normalmente el enlace entre los átomos de una celda es fuerte. Los átomos descolocados, empero, están sujetos débilmente y pueden moverse incluso con esfuerzos cortantes muy reducidos.

Cuando al hierro caliente se le agrega carbono, que luego precipita como cementita al enfriarse el metal, tiende a situarse en los espacios que proporcionan las dislocaciones en borde. Así inmoviliza éstas, disminuyendo la posibilidad de que un esfuerzo exterior pueda propagarlas a través de los cristales y provocar la rotura de los granos. Cuando Walton calienta su cuerda de piano y la deja enfriar por radiación y convección en el aire, los átomos de carbono disponen de tiempo para difundirse en el hierro y formar cementita en las dislocaciones. Con ello el alambre adquiere ductilidad suficiente para doblarse sin romperse.

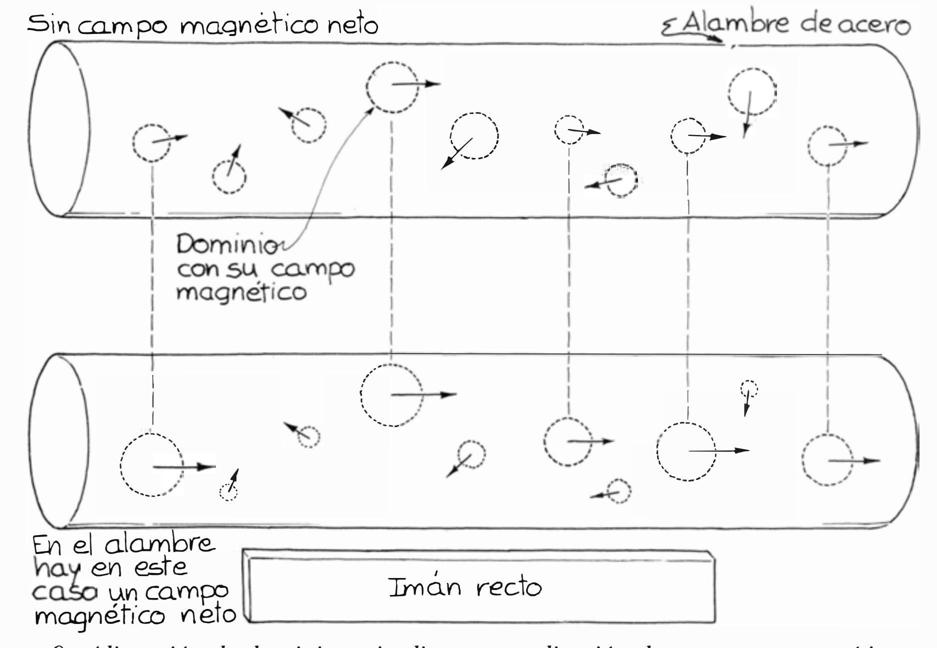
Cuando Walton calienta el alambre y lo enfría rápidamente con el agua de una esponja, obstaculiza la formación de cementita. Este temple sobreviene a tal velocidad que el carbono no puede difundirse. Y, aún más, sólo una parte del hierro gamma tiene tiempo para transformarse en hierro alfa. Aunque esta transformación es inicialmente rápida, enseguida se hace lenta. Por ello, la parte de alambre templada con agua se compone de una pequeña cantidad de hierro alfa y mucho hierro gamma. En ella todos los átomos de carbono quedan inmovilizados y resulta una solución sobresaturada de carbono, pues la concentración de éste en el hierro es superior a la normalmente posible a esa temperatura.

La formación de ésta nueva estructura, llamada martensita, no requiere ni difusión ni germinación. Se trata de un cambio espontáneo en la estructura cristalina del hierro gamma existente con átomos de carbono en algunas aristas. Aquí los átomos de hierro cambian de una estructura cúbica centrada en las caras a una estructura tetragonal centrada en el cuerpo. Como los átomos de carbono del hierro gamma no disponen de tiempo suficiente para difundirse, quedan atrapados en sus posiciones de las celdas unidad y fuerzan a los átomos de hierro de la estructura tetragonal centrada en el cuerpo a desplazarse de los puestos que les corresponden.

Esta deformación de los cristales crea zonas de gran tensión en la martensita. Esta deformación responde asimismo de la dureza de la martensita, ya que traba las dislocaciones en sus posiciones de los granos. Por esta razón, la martensita que crea Walton al enfriar rápidamente el alambre, cuando éste se encuentra caliente, resulta lo bastante dura para rayar el vidrio. Es asimismo frágil a causa de los numerosos puntos que sufren grandes tensiones dentro de los granos. Entonces, si el alambre se dobla, se rompe por culpa de esas tensiones internas.

El carbono de la martensita no permanece siempre atrapado, pero se difunde tan lentamente a través de los cristales que puede considerarse como si así fuera. Ahora bien, si aumentamos la temperatura, crece la velocidad de difusión, lo que permite que el carbono se reúna en minúsculas cantidades de cementita. Entonces, con temperatura y tiempo suficiente, el acero podrá recuperar más propiedades de la perlita.

El tercer misterio que suscita el trabajo de Walton atañe a la magnetización del alambre. Se dice que el hierro es ferromagnético. Una de las propiedades de este material se representa dividiéndolo en unas regiones llamadas dominios, cada uno de los cuales contribuye con un campo magnético en su vecindad. En conjunto, la sustancia puede parecer no magnética pues normalmente los campos magnéticos de los dominios se anulan unos con otros.



9. Alineación de dominios cristalinos por aplicación de un campo magnético

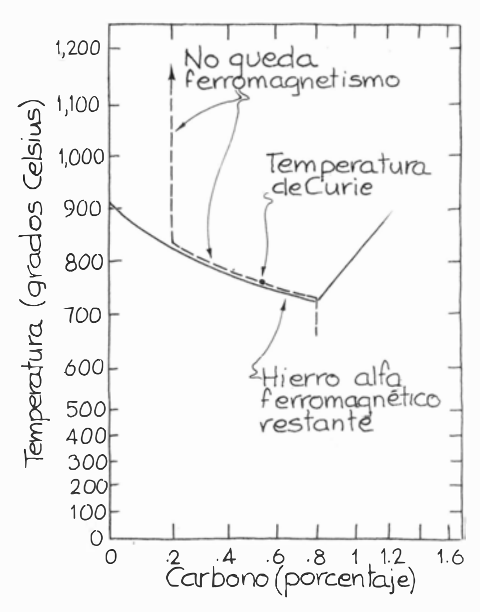
Tal es el caso de la cuerda de piano hasta que se le acerca un imán. Entonces, el campo del imán reorienta los dominios de modo que dotan al alambre de un campo magnético no nulo. Esta orientación se debe principalmente a variaciones en el tamaño de los dominios. Todo dominio cuyo campo magnético sea paralelo al del imán se alarga a expensas de los dominios adyacentes que tengan otras orientaciones.

El resultado de este crecimiento de los dominios es que el alambre y el imán se atraen. A la temperatura ambiente, tal atracción es suficientemente intensa para mover el alambre. A la temperatura característica del hierro gamma, la atracción está ausente.

Cuando se calienta el alambre, el aumento de temperatura provoca una mayor agitación de los átomos y moléculas. Dicha agitación comienza a perturbar la organización del campo magnético de cada dominio. Y cuanto más se calienta la sustancia, tanto más se debilita el campo magnético de cada dominio. Al final, la organización de los dominios desaparece por completo. Esto acontece a la llamada temperatura de Curie, bautizada así en honor de Pierre Curie, quien, en 1894, descubrió que el hierro pierde su ferromagnetismo cuando se calienta por encima de los 768 grados Celsius.

El origen esencial del campo magnético de las sustancias ferromagnéticas como el hierro no se conoce todavía. Parece ser que, cuando un átomo de hierro se une a un cristal, se ioniza parcialmente (pierde un electrón o más), pues los enlaces de los electrones de su órbita más externa se debilitan en presencia de otros iones de hierro. Estos electrones, si bien no están por completo libres, son lo bastante móviles para brincar de un ion a otro.

Todos los electrones tienen un campo magnético. Aunque se desconozca el origen del mismo, se trata de una característica del electrón, igual que lo es su carga eléctrica. Cuando los electrones saltan de un ion a otro dentro de un cristal, se influyen mutuamente merced a lo que se llama interacción de intercambio. Ello hace que los campos magnéticos de los electrones semilibres se alineen con una misma dirección. Por otra parte, un electrón que alineara su campo magnético en cualquier otra dirección necesitaría más energía. Así pues, la organización de los campos de los electrones de un dominio, la cual confiere al dominio un campo magnético neto, proviene de una minimización de la energía asociada a la interacción de intercambio.



10. Pérdida del ferromagnetismo

Conforme aumenta la temperatura, la agitación térmica destruye el comportamiento cooperativo de los electrones, hasta que, por encima de la temperatura de Curie, se desorganiza la estructura del dominio. A esta temperatura los electrones semilibres siguen teniendo su campo magnético cada uno, pero no organizados en conjunto. Si al hierro que esté más caliente que la temperatura de Curie se acerca un imán, apenas si se sentirá atraído por éste. El campo magnético del imán podrá orientar algunos de los campos de los electrones, pero esa alineación será fugaz por mor de la agitación térmica.

La pérdida del magnetismo puede seguirse en un diagrama de fases. Consideremos la perlita de un acero de carbono bajo. Cuando se calienta hasta el eutectoide y comienza a conformarse en hierro gamma y carbono, el hierro transformado pierde el ferromagnetismo. A la vez que el trayecto de calentamiento prosigue su ascenso a lo largo de A_3, el hierro alfa restante mantiene su magnetismo hasta que el metal alcanza el punto de Curie, a partir del cual parte alguna del hierro (ni siquiera el alfa) es ferromagnética.

Del efecto de la temperatura sobre la estructura cristalina del acero podría hablarse mucho más. Fotografiando al microscopio muestras atacadas químicamente puede estudiarse la precipitación del carburo de hierro en los aceros. Cuando éstos contienen más carbono que los que yo he tratado aquí, aparecen nuevas formaciones. Le dejo al lector el estudio de sus causas.

Libros

Expediciones botánicas del siglo XVIII, lógica, levaduras y fauna acuática de Centroamérica

Luis G. Ballester, J. de Lorenzo, R. Sentandreu y Angel Guerra

SPANISH SCIENTISTS IN THE NEW WORLD. THE EIGHTEENTH-CENTURY EXPEDITIONS, por Iris H. W. Engstrand. XIV + 220 pp., ilus., bibl., ind. Seattle-London, University of Washington Press, 1981. Desde que el Nuevo Mundo fue descubierto en 1492, la Corona de España mostró especial interés por conocer todos los aspectos de las nuevas tierras utilizando la ciencia como instrumento de conocimiento o, si se quiere, recurriendo a la razón humana como instrumento autónomo de conocimiento del mundo. Este interés se hizo más patente en dos momentos: durante los dos primeros tercios del siglo XVI y durante el siglo XVIII, especialmente en su segunda mitad. El Nuevo Mundo se convirtió para muchos científicos españoles del siglo XVI en objeto de estudios empíricos realizados con el criterio de comprobar directa y objetivamente los hechos. Las novedades del Nuevo Mundo se convirtieron, de acuerdo con la lograda expresión de I. B. Cohen (1960), en "source of science for Europe". Quizá la expedición científica española más madura y ambiciosa del siglo XVI fue la que el rey Felipe II encargó al médico y naturalista Francisco Hernández (1517-1587), que tenía por objeto estudiar la historia natural americana, y no sólo los aspectos medicinales, con una pauta metódica rigurosa. Por desgracia, el resultado de los trabajos de Hernández en Nueva España (1571-77) no se llegó a imprimir en el siglo XVI. Una serie de complejas circunstancias hizo que España olvidara esta tradición propia y, por otra parte, no participara en la llamada "revolución científica" en el siglo XVII y que se incorporara a ésta tardíamente.

Durante el siglo XVIII, España se incorporó a una actividad europea que hizo de la ciencia no sólo un método de conocimiento de la realidad mediante la observación y la experimentación, sino también una fuente de recursos con la que incrementar la riqueza de las naciones. Este proceso culminó en los reinados de Carlos III (1759-88) y de Carlos IV (1788-1808). En esos reinados alcanzaron su madurez instituciones científicas como el Jardín Botánico, la Real Academia de Medicina, el Observatorio astronómico, los Colegios de Cirugía, etcétera. Al mismo tiempo se establecieron estrechas relaciones con los científicos e instituciones europeas. Se pudo así volver a plantear, desde estos nuevos supuestos, el estudio de los productos naturales y las gentes de los dominios españoles en América y en el Pacífico; a la vez que reunir datos científicos sobre los que basar las disputas territoriales en tierras americanas con Inglaterra. Los científicos españoles del siglo XVIII que participaron en las expediciones científicas al Nuevo Mundo no practicaron el adanismo científico. Por el contrario, conectaron con su propia tradición científica del siglo XVI. A ello se debió que la expedición científica a Nueva España dirigida por el naturalista Martín de Sessé y Lacasta (1751-1808) tuviera como uno de sus objetivos "suplir, ilustrar y perfeccionar –así fueron sus palabras–, con arreglo al estado actual de las ciencias naturales, los escritos que dejó el Dr. Francisco Hernández".

El libro de Iris Engstrand describe dos de las expediciones científicas españolas del siglo XVIII: la primera, la Real Expedición Científica a Nueva España (1785-1800), que fue dirigida por el naturalista Martín de Sessé; la segunda (1789-94), por el marino y geógrafo Alejandro Malaspina (1754-1809). La autora maneja dos hipótesis. La primera es que "students of Europe's Age of Enlightenment often overlook the contributions of Spain in the realm of scientific achievement... the (Spanish) scientists who visited the Americas are scarcely known" (p. 3). La segunda, en relación causal con la anterior, es que "the results of their labors..., were not available to the European intellectual community" (*ibid.*). Y ello por varias razones. Por una parte, por lo que ella llama tradicional (!) uso del secreto por parte del gobierno español para negar a los enemigos el conocimiento científico y tecnológico; por otra, debido a las crisis políticas que afectaron a España en los años finales del siglo XVIII y comienzos del siglo XIX y que culminaron con la guerra contra Francia, y también guerra civil, entre 1808 y 1812. Hay una tercera: el insuficiente –por no decir nulo– desarrollo institucional que la historia de la ciencia ha tenido en España y en su universidad. Todo ello provocó el que los logros científicos de las expediciones se perdieran, se dispersaran por el mundo o permanecieran archivados, en lugar de publicarse. Ahora bien, la autora no puede olvidar que el secreto, relacionado con la actividad científica y tecnológica, fue práctica habitual en la Europa del siglo XVIII, practicándose por todas las naciones el llamado espionaje científico e industrial. Por desgracia, la libre comunicación de las ideas científicas y tecnológicas sigue hoy día sometida a intereses bastardos ajenos a los puramente científicos. El libro tampoco analiza en profundidad el cúmulo de circunstancias que confluyeron en España en la transición de los siglos XVIII al XIX, y que provocaron el que las empresas científicas –años antes estimuladas y pagadas– ahora se olvidaran y sus científicos quedaran parcial o totalmente marginados.

Un empeño, pues –que no deja de ser emocionante para el lector español–, que late a lo largo del libro es dar a conocer las contribuciones de España a la ciencia Europea del siglo XVIII mediante sus expediciones científicas al Nuevo Mundo y al Pacífico. Contribuciones que forman parte también de la historia de las jóvenes naciones americanas al Norte y al Sur del Río Grande. La expedición de Sessé fue una de las más importantes de la España ilustrada, junto con las dirigidas por José Celestino Mutis (1732-1808) a Nueva Granada, o las de Hipólito Ruiz (1752-1816) y José Pavón (1754-1840) al Perú. Bajo la dirección de Sessé y contando con el apoyo personal y científico de Casimiro Gómez Ortega, uno de los más importantes mentores de las iniciativas científicas en la España del momento y director del Jardín Botánico de Madrid, se fundó el Jardín Botánico en la ciudad de México, donde uno de los miembros de la expedición –Vicente Cervantes (1755-1829)– fue el primer profesor de botánica en México. La expedición recorrió los territorios de México, California, la costa noroeste del Pacífico, América Central (Guatemala), Puerto Rico y Cuba. En 1803, Sessé volvió a España con el resultado de su ingente trabajo en el que colaboraron botánicos, naturalistas, cirujanos, farmacéuticos, dibujantes y personal auxiliar. Por desgracia, sus manus-

critos quedaron inéditos. La expedición científica de Malaspina recorrió América del Sur, México, Alaska, la costa noroeste del Pacífico, California, las Filipinas, Australia y algunas islas del Pacífico. Tras una cuidada preparación en la que intervinieron las más prestigiosas instituciones científicas europeas (Collège de France, Royal Society) y españolas, se construyeron adrede dos corbetas dotadas de biblioteca y laboratorios con instrumentos astronómicos, náuticos, meteorológicos, físicos, químicos, biológicos y geodésicos. En ella intervinieron prestigiosos científicos españoles, nacidos en Latinoamérica o extranjeros al servicio de la Corona española como José Espinosa (1763-1815), Felipe Bauzá (1764-1834), D. Alcalá Galiano (1762-1805), Antonio Pineda (1753-92), Luis Neé (1734-1803), Tadeo Haenke (1761-1817) y un largo etcétera. Cuando Malaspina regresó en 1794, llevaba consigo un material que se encuentra entre los más ricos de las expediciones científicas europeas del siglo XVIII. Víctima de intrigas cortesanas, fue desterrado a sus posesiones de Italia, donde murió. Los materiales de la expedición se dispersaron.

El libro, al que precede un breve prefacio y un capítulo superficial sobre el "Spain's scientific Enlightenment" está dividido en 12 capítulos. Siete de ellos dedicados a la descripción de las actividades de Sessé y de los miembros de su expedición (capts. 2, 3, 8, 9, 10, 11 y 12) y los otros cuatro a la expedición de Malaspina (capts. 4, 5, 6 y 7). Lleva incorporados cuatro apéndices con listas de material científico y una amplia bibliografía.

El libro de Engstrand es descriptivo, a medio camino entre la investigación y la divulgación. Creo que el lector y la propia obra hubieran ganado si la autora hubiera introducido un capítulo historiográfico donde se analizaran los importantes estudios dedicados al tema y los supuestos desde los que están realizados: los trabajos de Fernández Duro (1876-81), H. W. Rickett (1947), Barras de Aragón (1950), Alvarez López (1950ss.), Jaramillo-Araujo (1952), Cutter (1960ss.), L. B. Simpson (1961), Bonifacio del Carril (1961), los de la propia autora (1963ss.), Steele (1964), Arias Divito (1968), R. McVaugh (1977), etcétera, algunos de los cuales la autora no recoge en su amplia bibliografía. A pesar de la rica documentación manejada, gran parte de la historia contada en esta obra ya es conocida por una persona mínimamente familiarizada con este capítulo de la historia de la ciencia española.

Creo que Engstrand no hace buen uso de los apéndices. Estos no deben incorporarse a un libro como mero ornamento. Deben formar parte integrante de la obra, bien como datos probatorios, material para ayudar al lector a contactar con la realidad estudiada o bien como *otro* modo de exposición; en cualquier caso, deben ofrecerse debidamente anotados. Por ejemplo, un listado de libros o de instrumental científico sólo adquiere sentido –especialmente cuando no se ha comentado ni utilizado de forma regular a lo largo de la exposición del libro– con las notas y comentarios historicocientíficos, sucintos, si se quiere, pero claros. No es el caso del libro que comentamos. Y es una pena, porque el material incorporado es muy interesante para el estudio de la penetración y difusión de las ideas científicas en la España ilustrada del período que se estudia. Este es, quizá, parte del precio que la autora paga por un libro que no posee un equilibrado balance entre la divulgación –alta divulgación– y la investigación.

El libro tiene algunos errores tipográficos: por ejemplo, 1820 por 1802 (p. 24, n. 36), *Desubierta* por *Descubierta* (p. 44, n. 1), Arcadio Pineda por Antonio Pineda (p. 143), eraduits por traduits (p. 186), Monveau por Morveau (p. 186). El libro está bellamente impreso con ilustraciones, grabados y mapas, algunos inéditos. (L. G.)

¿QUÉ ES LA LÓGICA MATEMÁTICA? por J. N. Crossley, C. J. Ash, C. J. Brickhill, J. C. S. Stilwell, N. H. W. Williams. Ed. Tecnos, M. 1983. Traducción de J. Alcolea Banegas. Constituye el resultado de unas conferencias pronunciadas bajo la dirección de Crossley en las universidades de Monash y Melbourne en 1971 con el objetivo de popularizar las nociones básicas de la lógica matemática, hacerlas asequibles al gran público. Divulgación que se quiere ligada, en todos los casos, a la supresión del instrumental matemático que se exige de los interesados profesionalmente por la lógica. Y es un primer punto que cabría tocar: si es posible tal popularización sin pérdida no ya del rigor, sino de todo un haz de precisiones que exigen de la formulación matemática, que si se ha utilizado no es, precisamente, por capricho de algunos que se autotitulen 'lógicos'. No se trata de detenerse en el puro formulismo, sino en las ideas subyacentes.

En la presentación, Valdés trata sobre "La idea de infinito y la lógica matemática" con un breve, y acertado, recorrido por la problemática que la aceptación del infinito actual provocó en el hacer matemático con la teoría de conjuntos de Cantor y las paradojas o antinomias asociadas, así como su repercusión en la formalización de las teorías matemáticas y, por consiguiente, en la lógica matemática, con la aparición de dos grandes corrientes: la teoría de modelos y la recursividad. Es una presentación que complementa el primer capítulo en el que se esboza un panorama histórico de la lógica. Panorama cuyo mayor valor se centra en insistir en el reduccionismo de la matemática a un lenguaje simbólico y en los problemas que surgen de este programa que, por supuesto, no es del ámbito lógico sino del ámbito de las creencias. Problemas que dan origen a las nociones de completitud, decisión y computabilidad, categoricidad... Y a tratar de la completitud del cálculo de predicados se dedica el segundo capítulo. Y es donde la popularización, a veces, muestra sus peligros como en el no distinguir los matices del término o la dificultad de dar una idea clara de la noción de interpretación de un lenguaje en una estructura dada. Salvadas algunas de tales imprecisiones por las notas de Valdés, el capítulo ofrece un esquema de la línea de construcción de modelos sugerida por Henkin y que supone la construcción de extensiones maximalmente consistentes; esquema que puede ser útil para quien, posteriormente, dedique su tiempo a estudiar lógica, aunque quizá no se aclare el por qué se requieren las condiciones de maximalidad al agregar las constantes "testigo" o el por qué de las mismas.

Ciertamente ello se complementa en el capítulo siguiente, que esboza alguno de los temas centrales de una de las divisiones de la lógica: la teoría de modelos. Aquí se enuncian dos de sus resultados fundamentales: el teorema de compacticidad, cuyo esbozo de demostración se apoya en la completitud, y el teorema descendente de Löwenheim-Skolem-Tarski. Y el capítulo cuarto se dedica a la otra gran división de la lógica: la calculabilidad. Tema que se introduce a partir de la noción de computabilidad mediante las máquinas de Turing, aceptando sin más discusiones la tesis de Church. Se pasa a la noción de función parcialmente recursiva y, a su través, se intenta dar cuenta de uno de los argumentos clave en los razonamientos de Cantor, el de la diagonal, al que se achacó ser impredicativo o autoreferencial y, por ello, base de la aparición de las antinomias conjuntistas.

Es interesante señalar que los autores abordan el problema de la máquina universal, así como el de palabras, para llegar a la afirmación de la indecidibilidad del cálculo de predicados. Ello

conduce, naturalmente, a las cuestiones de incompletitud, que constituyen uno de los resultados más espectaculares asociados al nombre de Gödel, respecto de los sistemas formales que sean lo suficientemente potentes como para incluir las nociones de aritmética. Y estas son las cuestiones que se plantean en el capítulo quinto, donde se esboza el proceso de aritmetización de Gödel, lo que hoy se conoce en algunos medios con el término de gödelización, para codificar las fórmulas, las sucesiones de fórmulas de un sistema formal. Hasta aquí, tanto en la presentación como en los cinco primeros capítulos, se ha esbozado alguno de los temas centrales de la lógica como pudieran haber quedado en los años cuarenta. El libro termina con un capítulo dedicado a la teoría de conjuntos, donde, tras una introducción amena, se pasa a uno de los resultados más avanzados –en cuanto a época, me refiero–: el método de forcing de Cohen, de 1963, y con él la aparición de conjuntos genéricos o forzados a cumplir unas condiciones determinadas. Método con el cual se pretende dar una idea de que la hipótesis del continuo es independiente de la teoría de conjuntos en la formulación ZFS.

Este libro, como introducción, y tal como se muestra en la versión castellana, puede cumplir un digno papel. Con algunas advertencias. La popularización planteada como objetivo exige suprimir, es lo que afirman los autores, el instrumental matemático. Naturalmente a los propios autores les es imposible la total supresión simbólica. De aquí que el libro se quede a medio camino. Mayor obstáculo es la no advertencia al lector del hecho de que lo expuesto no constituye más que una parte, ya muy clásica, de los temas de que trata la lógica. Si se habla de la importancia de los teoremas de compacticidad y Löwenheim-Skolem-Tarski, no se menciona que los resultados de Lindstrom de 1968 indican que son las dos notas caracterizadoras, precisamente, de la lógica de orden uno. Ni se mencionan las lógicas o sistemas formales con otros tipos de cuantificación o de longitud infinita, ni si la consistencia y con ella los temas relacionados van referidos a un tipo especial de conectiva, la negación exclusiva, así como no hay referencia a sistemas formales alternativos a los aquí expuestos, exclusivamente veritativo-funcionales y, además, bivalentes... La lógica, hoy, posee núcleos y temas más amplios que los aquí contenidos.

En cuanto a la traducción se plantean pequeños problemas, ya antiguos y en los que seguimos sin acuerdo, respecto a la terminología. Así, un espacio compacto es el que satisface la condición del subrecubrimiento finito. Esta condición es la que los matemáticos califican como teorema de compacticidad desde hace ya tres cuartos de siglo. Y de esa matemática lo adoptó Malcev a la lógica. Los autores españoles, si matemáticos, mantienen ese término: "compacticidad" en paralelo a la compacticidad que le dio nombre; si no matemáticos, traducen esa condición con el término "compacidad" que es como aparece en este libro; análogo respecto a términos como los de "completitud" frente a "completud", o "terna ordenada" frente a "triplo"... (J. L.)

YEASTS: CHARACTERISTICS AND IDENTIFICATION, por J. A. Barnett, R. W. Payne y D. Yarrow. Cambridge University Press, 1983. Las levaduras constituyen un grupo heterogéneo de hongos que se distinguen de los filamentosos en que su forma dominante es unicelular y su reproducción es asexual, por gemación. Este grupo de microorganismos, que se encuentra estrechamente vinculado con el bienestar del hombre, ha sido estudiado desde el punto de vista de su identificación, nomenclatura y clasificación de especies, fundamentalmente por los investigadores del Centraalbureau voor Schimmelcultures de Delft. Uno de ellos, Jacomina Lodder, publicó en 1970 un tratado: *The Yeast: A taxonomic study* (North-Holland, 1970) que se convirtió en la obra de obligada referencia de todos los interesados en este campo. Describía las características de 349 especies de levaduras poniendo un énfasis especial en su capacidad de formar ascosporas. Esta característica plantea al investigador, en ocasiones, inconvenientes de manejo insuperables.

Dado el interés que hoy despierta este grupo de microorganismos, no se olvide que algunas de sus especies se utilizan en el estudio de problemas básicos de la biología de los seres superiores al objeto de obtener con ellos resultados espectaculares en la utilización de las nuevas tecnologías derivadas de la genética molecular, reviste enorme importancia la aparición del volumen que reseñamos, verdadero hito en la técnica empleada en la identificación de las levaduras.

La finalidad de la obra es doble: por un lado, facilitar el uso de las levaduras, tanto en los laboratorios industriales como en los de investigación básica, al proporcionar una información actualizada de las 473 especies incluidas. Por otro lado, facilita el trabajo de identificación. A tal fin no sólo ofrece el aspecto microscópico de cada especie, sino también la descripción sistemática de las mismas mediante una combinación de claves y tablas.

En la introducción se indican algunos aspectos que pueden ayudar a encontrar nuevos usos de las levaduras. Hasta hoy, la industria e investigación recurren de una manera casi exclusiva a *Saccharomyces cerevisiae*, y ya es tiempo de que sean utilizadas otras especies más versátiles en cometidos específicos. Además, esta versatilidad puede incrementarse mediante la transferencia de genes y la fusión de protoplastos. La explotación de nuevas especies de levaduras depende de los conocimientos que tengamos de sus propiedades, labor acometida únicamente por los taxónomos.

El capítulo 2 expone la realización de pruebas fisiológicas y toma de microfotografías. De las primeras se ocupó uno de los autores (Yarrow). Los resultados obtenidos con las distintas especies son aquí plenamente comparables, cosa que no ocurre cuando abrimos otros textos en los que los resultados han sido obtenidos por varios científicos en distintos laboratorios y quizás utilizando técnicas diferentes. Las condiciones en que se realizaron las pruebas fisiológicas quedan suficientemente definidas, lo que permite repetirlas con garantía de éxito. La obtención de preparaciones para la observación microscópica de las levaduras queda también perfectamente explicada y, en conjunto, podemos decir que son de una calidad excelente. Un inconveniente, sin embargo: la observación de las preparaciones se realizó con microscopía óptica, no con microscopía electrónica, lo que empaña un tanto el valor de las mismas.

El capítulo 3, dedicado a la clasificación, justifica las características escogidas: apariencia microscópica de las células, forma de reproducción sexual, propiedades fisiológicas (fundamentalmente las nutricionales) y peculiaridades bioquímicas.

La denominación de las distintas especies de levaduras se encuentra regulada por el Código Internacional de Nomenclatura Botánica. (A pesar de ello, todavía se leen nombres que no concuerdan con dicha normativa legal.) Destacaríamos aquí también la atención prestada a los problemas relativos a la determinación de las levaduras, que se derivan de la utilización de la taxonomía numérica o de la complementariedad de las secuencias de ADN. En el capítulo 4 se reseñan los métodos de laboratorio utilizados en la identificación de las levaduras. Entre ellos, la apariencia microscópica de las células

vegetativas y ascosporas o balistoconidios, la habilidad de utilizar compuestos orgánicos como única fuente de carbono o de nitrógeno, la habilidad de ciertos azúcares en condiciones anaeróbicas y otros.

Sigue un resumen de las características de los distintos generos de levaduras, para ampliarlo, en el capítulo sexto, con una tabulación de las características de las especies incluidas. En el capítulo 7, el más extenso de la obra, se recoge la descripción y microfotografía de cada una de las especies reconocidas por el Centraalbureau voor Schimmelcultures hasta enero de 1983. Los dos capítulos siguientes, el 8 y el 9, comprenden las claves y tablas utilizadas en la identificación de las distintas especies, mientras que en el último se recogen todos los nombres y sinónimos, junto con la referencia al autor que usara primeramente cada denominación y de los que emplearon otras luego. (R. S.)

AQUATIC BIOTA OF MEXICO, CENTRAL AMERICA AND THE WEST INDIES, por Stuart H. Hurlbert y Alejandro Villalobos-Figueroa (eds.). Universidad Autónoma Metropolitana de México, 1982. La historia de esta obra comenzó en forma de un proyecto que informalmente se denominaba *Aquatic Biota of Latin America*. Este libro es el tercero y último de la serie. Los dos anteriores se publicaron en 1977 y 1981, respectivamente.

El objetivo primordial de estos tres manuales es proporcionar a especialistas de diferentes campos de la biología un acceso cómodo a la bibliografía existente sobre la flora y la fauna de las aguas continentales de Latinoamérica. Información que es bastante amplia y estaba muy dispersa. Con estas publicaciones se pretende, además, promocionar la investigación taxonómica y ecológica en la citada área geográfica.

El libro está concebido de manera que cada recopilación bibliográfica vaya precedida de una introducción realizada por especialistas en cada uno de los grupos taxonómicos abordados. Esta introducción está escrita en inglés y en español. Los autores del volumen son 79: 49 norteamericanos, 12 mexicanos, cuatro canadienses, cuatro brasileños, tres alemanes, dos australianos, dos ingleses, un francés, un venezolano y un argentino. Y en él se estudian 59 taxones, que van desde las bacterias y los hongos hasta los mamíferos. Son 529 páginas de apretada información. La edición se ha hecho en rústica y lleva en la portada un buen dibujo de *Atya scabra*, crustáceo decápodo de distribución pantropical que se ha tomado como símbolo.

La recensión de una obra de estas características no es tarea fácil por su falta de unidad y por la extrema variedad de la información que contiene. Como cualquier libro que carezca de una ilación o cuerpo doctrinal únicamente informa pero no forma. Se trata de un laborioso y bien presentado compendio que, a primera vista, es útil para el investigador. Y decimos a primera vista porque, para saber con certeza el servicio que puede prestar hay que conocer la calidad, actualidad y acierto de la selección bibliográfica que recoge. Como estar al día en todos los grupos taxonómicos sobre los que versa este libro es imposible para una persona sola, hemos optado por hacer unas cuantas catas revisando aquellos grupos con los que estamos más familiarizados. Así, en los moluscos y en los peces la selección bibliográfica parece bastante cuidada y completa, habiéndose encargado a especialistas competentes. Otros taxones están menos trabajados, por lo que el nivel del texto es desigual.

Llama la atención la pequeña proporción de científicos nativos que han colaborado, así como el relativo escaso número de citas de trabajos realizados por investigadores de los países de las áreas involucradas. Está, por tanto, bien justificada la llamada de ánimo –casi grito– que los editores lanzan en la introducción para que aumente el número de científicos, sobre todo hispanoparlantes añadimos nosotros, que se dediquen a la taxonomía.

Taxonomía. Hemos topado aquí con un término actualmente controvertido. Para algunos biólogos, es disciplina en decadencia: un campo que tuvo su esplendor en épocas pretéritas, pero que ya no tiene vigencia o a la que le resta poco por hacer. Para otros, esa actitud es desacertada y miope. Piensan estos últimos que aunque se deba evitar la manera de enfocar la taxonomía de ciertos autores del pasado, otros la hicieron de tal forma que sus trabajos son auténticas joyas científicas. Afirman que, evitando caer en desenfoques que llevan a querer describir una especie nueva en cuanto uno, o varios individuos, presentan ciertas peculiaridades con respecto a la media poblacional, y contrastando los estudios morfológicos y anatómicos con las técnicas de bioquímica, citotaxonomía, taxonomía numérica, etcétera, que se han desarrollado, la taxonomía no está obsoleta, ni carece de rigor, ni es superflua.

Lo que resulta innegable es que no da lo mismo que en un ecosistema esté presente una especie u otra. Y ello porque los seres vivos se comportan de forma diferente y más compleja que las partículas de materia inorgánica o que las moléculas. Son organismos, y como tales, estructuras materiales que gozan de la propiedad de constituir unidades funcionales cuyas partes cooperan al buen funcionamiento del todo. Por ello, sus manifestaciones no pueden comprenderse por completo a través del análisis químicos, físicos o morfológicos de sus elementos, mas tampoco pueden concebirse fuera de la materia, aunque no sean exclusivamente eso. Hay, por ejemplo, pautas de conducta que ligan a los individuos con sus propios congéneres, con los de otras especies y con el medio ambiente, que son particulares y específicas, y que no se explican puramente por un conjunto de reacciones fisicoquímicas encadenadas.

Por lo expuesto, y por otras razones, no se puede prescindir de la taxonomía a la hora de hacer biología. La experiencia enseña que hay determinadas ramas de la misma en que es imprescindible. Así, la mayoría de los ecólogos competentes de nuestro tiempo son buenos sistemáticos generales, y muy buenos en algún taxón particular. Esto no va en detrimento de que para la comprensión de los principios generales que rigen la transferencia de energía de un nivel a otro del ecosistema se apliquen modelos termodinámicos, o que las relaciones depredador-presa se intenten cuantificar mediante ecuaciones diferenciales. El problema no está en los modelos, sino en quienes creen en ellos a pies juntillas. Es intelectualmente sano y sensato conservar cierto escepticismo ante cualquier modelo, porque son sólo simplificaciones, a veces burdas, de la realidad.

El libro que estamos comentando, y los otros dos de la trilogía, no han sido casos aislados. En los Estados Unidos han aparecido recientemente varios tratados sobre organismos, de diferentes ecosistemas, lo cual demuestra que en ese país el interés por la taxonomía no ha decaído. Con lo desarrolladas que están las técnicas de manejo de ficheros mediante ordenador se podría haber hecho listados bibliográficos mucho más completos, pero esas listas habrían carecido del valor que tienen las que se presentan en este libro.

Por último, pensamos que la iniciativa de los editores de este compendio, además de útil, marca una pauta que deberíamos imitar en nuestro país, con la intención de preparar una flora y fauna de la Península Ibérica que no está hecha. Para realizar esta tarea quedan todavía huecos que llenar, pero hay ya mucho avanzado. (A. G.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

MAQUINAS DE TURING

ON NONCOMPUTABLE FUNCTIONS. Tibor Rado in *Bell System Technical Journal*, vol. 41, n.° 3, págs. 877-884; mayo, 1962.

THE UNDECIDABLE. Martin Davis. Raven Press, 1965.

INCOMPUTABILITY. C. A. R. Hoare y D. C. S. Allison en *Computing Surveys*, vol 4, n.° 3, págs. 169-178; septiembre, 1972.

ALAN TURING: THE ENIGMA. Andrew Hodges. Simon and Schuster, 1983.

ZONAS OCEANICAS DE FRACTURA

TOPOGRAPHY AND TECTONICS AT THE INTERSECTIONS OF FRACTURE ZONES WITH CENTRAL RIFTS. Norman H. Sleep y Shawn Biehler en *Journal of Geophysical Research*, vol. 75, n.° 14, págs. 2748-2752; 10 de mayo de 1970.

VERTICAL TECTONISM IN OCEANIC FRACTURE ZONES. Enrico Bonatti en *Earth and Planetary Science Letters*, volumen 37, número 3, págs. 369-379; enero, 1978.

MEMBRANE STRESSES NEAR MID-OCEAN RIDGE-TRANSFORM INTERSECTIONS. Kazuya Fujita y Norman H. Sleep en *Tectonophysics*, vol. 50, págs. 207-221; 20 de septiembre de 1978.

FORMERLY EMERGING CRUSTAL BLOCKS IN THE EQUATORIAL ATLANTIC. Enrico Bonatti y Andy Chermak en *Tectonophysics*, volumen 72, números 3-4, páginas 165-180; 10 de febrero de 1981.

ENDOCITOSIS DE PROTEINAS Y PARTICULAS

INTRACELLULAR SITE OF ASIALOGLYCOPROTEIN RECEPTOR-LIGAND UNCOUPLING: DOUBLE-LABEL IMMUNOELECTRON MICROSCOPY DURING RECEPTOR-MEDIATED ENDOCYTOSIS. Hans J. Geuze, Jan Willem Slot, Ger J. A. M. Strous, Harvey F. Lodish y Alan L. Schwartz en *Cell*, vol. 32, n.° 1, págs. 277-287; enero, 1983.

*p*H AND THE RECYCLING OF TRANSFERRIN DURING RECEPTOR-MEDIATED ENDOCYTOSIS. Alice Dautry-Varsat, Aaron Ciechanover y Harver F. Lodish en *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, n.° 8, págs. 2258-2262; abril, 1983.

MINIACERIAS

TECHNOLOGY AND STEEL INDUSTRY COMPETITIVENESS. Office of Technology Assessment, U.S. Government Printing Office, 1980.

THE STEEL INDUSTRY. W. Adams y H. Muller en *The Structure of American Industry*, dirigido por Walter Adams. Macmillan Publishing Co., 1982.

STEEL —UPHEAVAL IN A BASIC INDUSTRY. D. F. Barnett y L. Schorsch. Ballinger Publishing Co., 1983.

PARASITOS QUE CAMBIAN EL COMPORTAMIENTO DE SU PATRON

ALTERED COLOR, BEHAVIOR AND PREDATION SUSCEPTIBILITY OF THE ISOPOD *ASELLUS INTERMEDIUS* INFECTED WITH *ACANTHOCEPHALUS DIRUS*. Joseph W. Camp y Harry W. Huizinga en *The Journal of Parasitology*, vol. 65, n.° 4, págs. 667-669; agosto, 1979.

THE LIFE CYCLE OF *OCTOSPINIFEROIDES CHANDLERI* BULLOCK, 1957 (ACANTHOCEPHALA: NEOECHINORHYNCHIDAE), WITH SOME OBSERVATIONS ON PARASITE-INDUCED, PHOTOPHILIC BEHAVIOR IN OSTRACODS. David J. DeMont y Kenneth C. Corkum en *The Journal of Parasitology*, vol. 68, n.° 1, págs. 125-130; febrero, 1982.

ALTERED BEHAVIORAL RESPONSES IN INTERMEDIATE HOSTS—AN ACANTHOCEPHALAN PARASITE STRATEGY. Janice Moore en *American Naturalist*, vol. 123, n.° 4, págs. 572-577; abril, 1984.

EL UNIVERSO INFLACIONARIO

A NEW INFLATIONARY UNIVERSE SCENARIO: A POSSIBLE SOLUTION OF THE HORIZON, FLATNESS, HOMOGENEITY, ISOTROPY AND PRIMORDIAL MONOPOLE PROBLEMS. A. D. Linde en *Physics Letters*, vol. 108B, n.° 6, págs. 389-393; 4 de febrero de 1982.

COSMOLOGY FOR GRAND UNIFIED THEORIES WITH RADIATIVELY INDUCED SYMMETRY BREAKING. Andreas Albrecht y Paul J. Steinhardt en *Physical Review Letters*, vol. 48, n.° 17, págs. 1220-1223; abril, 1982.

UN CAMPAMENTO DE CAZADORES INDIOS DURANTE 20.000 AÑOS

MULTIDISCIPLINARY RESEARCH IN THE NORTHEAST: ONE VIEW FROM MEADOWCROFT ROCKSHELTER. J. M. Adovasio en *Pennsylvania Archaeologist*, vol. 52, n.os 3-4, págs. 57-68; septiembre, 1982.

EVIDENCE FROM MEADOWCROFT ROCKSHELTER. J. M. Adovasio, J. Donahue, K. Cushman, R. C. Carlisle, R. Stuckenrath, J. D. Gunn y W. C. Johnson en *Early Man in the New World*, dirigido por Richard Shutler, Jr. Sage Publications, Incorporated, 1983.

GANGLIOSIDOS

GLYCOSPHINGOLIPIDS: STRUCTURE, BIOLOGICAL SOURCE, AND PROPERTIES. B. A. Macher & Ch. C. Sweeley, en *Meth. Enzymol.*, volumen 50, páginas 236-251; dirigido por V. Ginsburg. Academic Press; Nueva York, 1978.

STRUCTURE AND FUNCTION OF GANGLIOSIDES. Dirigido por L. Svennerholm, P. Mandel, H. Dreyfus y P.-F. Urban en *Adv. Exp. Med. & Biol.*, vol. 125. Plenum Press; Nueva York, 1980.

GYCOSPHINGOLIPIDS IN CELLULAR INTERACTION, DIFFERENTIATION, AND ONCOGENESIS. S. Hakomori en *Ann. Rev. Biochem.*, vol. 50. Dirigido por E. E. Snell, P. D. Bayer, A. Meister y C. C. Richardson, Ann. Rev. Inc.; Palo Alto, California, 1981.

POSSIBLE FUNCTIONAL ROLE OF BRAIN GANGLIOSIDES IN ADAPTATIVE NEURONAL PROCESSES. H. Rahmann, en *Neuronal Transmission, Learning and Memory*. Dirigido por R. Caputto y C. Ajmone. Raven Press; Nueva York, págs. 159-177, 1983.

JUEGOS DE ORDENADOR

COMPUTER RECREATIONS: "DARWIN". Aleph Null en *Software: Practice & Experience*, vol. 2, n.° 1, págs. 93-96; enero-marzo, 1972.

68000 ASSEMBLY LANGUAGE PROGRAMMING. Gerry Kane y Doug Hawkins Legenthal. Osborne-McGraw-Hill, 1981.

TALLER Y LABORATORIO

STRUCTURE AND PROPERTIES OF ALLOYS. Robert M. Brick, Robert B. Gordon y Arthur Phillips. McGraw-Hill Book Company, 1965.

MAGNETISM IN SOLIDS. D. H. Martin. The MIT Press, 1967.